

## NEAR ROADWAY AIR QUALITY: A META-ANALYSIS

Contract AQ-04-01: Developing Effective and Quantifiable  
Air Quality Mitigation Measures

January 21, 2009

By

Alex Karner

Douglas Eisinger, PhD

Deb Niemeier, PhD, PE

### Abstract

**Background:** Interest in roadside air quality has been steadily increasing since the late 1970s, motivated by a growing body of published work that has investigated links between elevated concentrations of mobile source pollutants and potential adverse health impacts for individuals in areas near roadways. This study involved creating and then assessing a dataset to synthesize data from the literature on near roadway air pollution. The study goal was to improve understanding about the atmospheric transport and fate of air pollution emitted from vehicles on a roadway and carried downwind.

**Methods:** Using more than 30 studies from the peer-reviewed literature, we compiled a database linking pollutant concentrations to various distances from the road. We normalized concentrations across and within pollutant categories, identified which concentrations varied as a function of distance from road, and fit trend lines through the data to enable improved understanding of the concentration-distance relationship. We also tested whether statistically significant differences existed among concentrations measured at various distance ranges (0–80 m, 80–120 m, and > 120 m from the road) using an analysis of variance model.

**Results:** In total, we identified 572 measurements of pollutant concentrations and their associated distances from the road. Results of two statistical analyses (local regression and analysis of variance) showed broadly consistent results: carbon monoxide, elemental and black carbon, metal deposition, oxides of nitrogen, non-alkane hydrocarbons, and ultrafine particle number are consistently elevated near roads, reaching 1.7–20 times above background at the edge of road, depending on pollutant and normalization method. In general, pollutant concentrations approach background by approximately 75–400 m downwind of a road, depending on the pollutant; most pollutants decay to background by 150 m from the edge of road.

## ***About The U.C. Davis-Caltrans Air Quality Project***

<http://AQP.engr.ucdavis.edu/>

Department of Civil & Environmental Engineering  
University of California  
One Shields Ave., Davis, CA 95616  
(530) 752-0586

**Mission:** The Air Quality Project (AQP) seeks to advance understanding of transportation related air quality problems, develop advanced modeling and analysis capability within the transportation and air quality planning community, and foster collaboration among agencies to improve mobility and achieve air quality goals.

**History:** Since the 1990s, the U.S. Federal Highway Administration and Caltrans have funded the AQP to provide transportation-related air quality support. Caltrans and AQP researchers identify and resolve issues that could slow clean air progress and transportation improvements.

**Accessibility:** AQP written materials and software tools are distributed through our website, peer-reviewed publications, conference presentations, training classes, formal reports and technical memoranda, and periodic newsletters.

**Research:** AQP investigations focus on project-level, regional-scale, and national-level assessments. Tools and publication topics cover pollutant-specific problems such as those involving particulate matter, carbon monoxide, carbon dioxide, ozone and air toxics; activity data collection and assessment for on- and off-road vehicles and equipment; mitigation options such as transportation control measures; policy analyses addressing transportation conformity and state implementation plan development; litigation support; and goods movement assessments.

### **Project Management**

Principal Investigator and Director: Deb Niemeier, PhD, PE  
Program Manager: Douglas Eisinger, PhD

Caltrans Project Manager: Mike Brady, Senior Environmental Planner  
Air Quality and Conformity Coordination  
Division of Transportation Planning, MS-32  
California Department of Transportation  
1120 N Street, Sacramento, CA 94274  
(916) 653-0158

## **Acknowledgements**

Support for this research was provided by the UC Davis-Caltrans Air Quality Project and a Postgraduate Scholarship-Masters from the Natural Sciences and Engineering Research Council of Canada. Dr. Song Bai provided valuable statistical assistance.

## Table of Contents

<b>Acknowledgements .....</b>	<b>i</b>
<b>Table of Contents .....</b>	<b>ii</b>
<b>List of Figures.....</b>	<b>iii</b>
<b>List of Tables .....</b>	<b>iv</b>
<b>List of Acronyms .....</b>	<b>v</b>
<b>1 INTRODUCTION.....</b>	<b>1</b>
<b>2 METHODS OVERVIEW .....</b>	<b>4</b>
<b>3 LITERATURE SUMMARY .....</b>	<b>7</b>
<b>4 DATA SELECTION AND PREPARATION.....</b>	<b>9</b>
<b>5 RESULTS AND DISCUSSION .....</b>	<b>13</b>
5.1 Unnormalized near roadway concentrations.....	13
5.2 Normalized near roadway concentrations.....	16
5.2.1 Normalization methods.....	16
5.2.1.1 Overview Discussion: Normalized to Background.....	17
5.2.1.2 Overview Discussion: Normalized to Edge.....	18
5.2.2 Results: Near roadway concentrations normalized to background values.....	21
5.2.2.1 Trend analysis .....	22
5.2.2.1.1 Local regression .....	23
5.2.2.1.2 Analysis of variance: Overview Discussion .....	33
5.2.2.1.3 Analysis of variance: Results .....	37
5.2.3 Results: Near roadway concentrations normalized to roadway edge .....	39
5.2.3.1 Trend analysis .....	41
5.2.3.1.1 Local regression .....	41
5.2.3.1.2 Analysis of variance.....	46
5.2.4 Comparison of normalization methods.....	48
<b>6 LIMITATIONS AND FUTURE WORK.....</b>	<b>51</b>
<b>7 CONCLUSIONS .....</b>	<b>53</b>
<b>REFERENCES.....</b>	<b>55</b>
<b>Appendix A: R code.....</b>	<b>64</b>
<b>Appendix B: Review of the near road pollution concentration literature.....</b>	<b>84</b>
<b>Appendix C: Database.....</b>	<b>104</b>

## List of Figures

Figure 1: Summarizing the data: (a) histogram of all observations grouped by 50 m distance bin (b) barchart showing number of studies grouped by pollutant type. ....	8
Figure 2: All collected near road air pollution data plotted in separate panels.....	15
Figure 3: Background normalized near road air pollution data. ....	22
Figure 4: Local regression of background normalized concentrations on distance (a) for the entire range of data; (b) for a subset of the data limited to five times above background concentrations, to show increased detail. ....	26
Figure 5: Background normalized near road air pollution data plotted in separate panels.....	32
Figure 6: Edge normalized near road air pollution data. ....	40
Figure 7: Local regression of edge normalized concentrations on distance. ....	43
Figure 8: Edge normalized near road air pollution data plotted in separate panels. ....	45

## **List of Tables**

Table 1: Summary of Observations .....	7
Table 2: Particle Size Ranges Counted in Studies Which Recorded Particle Numbers .....	13
Table 3: Summary Statistics Calculated Over the Entire Distance Range Grouped by Pollutant Type .....	16
Table 4: ANOVA Results for Background Normalized Pollutant Concentrations .....	38
Table 5: ANOVA Results for Edge Normalized Pollutant Concentrations.....	48
Table 6: Summary of Pollutant Profiles Under Both Normalization Methods.....	50

## **List of Acronyms**

ADT	Average daily traffic
ANOVA	Analysis of variance
BN	Background normalization
CARB	California Air Resources Board
CO	Carbon monoxide
CO <sub>2</sub>	Carbon dioxide
CPC	Condensation particle counter
EN	Edge normalization
EPA	United States Environmental Protection Agency
GIS	Geographic information system
I-5	Interstate 5
NAAQS	National Ambient Air Quality Standards
NASSCO	National Steel and Shipbuilding Company
NO	Nitric oxide
NO <sub>2</sub>	Nitrogen dioxide
NO <sub>x</sub>	Total oxides of nitrogen
PAH	Polycyclic aromatic hydrocarbons
PM	Particulate matter
PM <sub>10</sub>	Particulate matter less than 10 µm in aerodynamic diameter
PM <sub>2.5</sub>	Particulate matter less than 2.5 µm in aerodynamic diameter
SMPS	Scanning mobility particle sizer
SO <sub>2</sub>	Sulfur dioxide
TAC	Technical Advisory Committee
TSP	Total suspended particles
VOC	Volatile organic compound

# 1 INTRODUCTION

Interest in roadside air quality has been steadily increasing since the late 1970s, most recently driven by a growing body of published work that has investigated links between elevated concentrations of ultrafine particles and other mobile source pollutants and potential adverse health impacts for individuals in areas near roadways. The near road interest marks a progression from earlier studies which focused on regional-level pollution and had confirmed the link between air pollution at that scale and public health (see, e.g., Dockery et al. 1993; or the review by Brunekreef and Holgate 2002). The roadside-health link has been strengthened by a number of public health studies conducted beginning in the early 1990s to assess links between illness and exposure to roadside pollutants, usually measured by a proxy indicator of automobile traffic such as distance from road or traffic density within a buffer zone. Some of these studies have found associations between illnesses and exposure to traffic or traffic-related proxy measurements (such as residential distance to roadway) including asthma (Edwards et al. 1994; Lin et al. 2002; Kim et al. 2004; Zmirou et al. 2004; Gauderman et al. 2005; McConnell et al. 2006), cancer (Savitz and Feingold 1989), allergic sensitization (Wyller et al. 2000), heart disease (Peters et al. 2004; Brugge et al. 2007), and even schizophrenia (Pedersen et al. 2004). Other studies have found no significant links between exposure to automobiles and adverse health (Wjst et al. 1993) or found mixed results (Zmirou et al. 2004).

These inconsistencies arise in part due to the imprecise nature of the proxy measurements used to estimate pollutant exposure—relying on self-reported traffic exposure (see, e.g., Behrens et al. 2004) or various other traffic density metrics<sup>1</sup> coded from geographic information system (GIS) data based on the location of primary residence combined with self-reported incidents of respiratory symptoms (see, e.g., Duhme et al. 1996) or children's symptoms (see, e.g., Weiland et al. 1994), among others. However, as individuals go about their day, exposures are not uniformly distributed temporally or spatially (Sexton and Ryan 1988), so using the primary residence as the

---

<sup>1</sup> For example, Murakami et al. (1990) simply defined distance to “main busy roads,” Reynolds et al. (2004) define road density as the length of road contained within a 500 ft area around the home and traffic density as VMT/mi<sup>2</sup> again within a 500 ft circular area around the home, while Lin et al. (2002) use distance to state route, heavy trucks passing within 200 m and 500 m of the home and traffic density (VMT) within the same buffer zones. Other proxies can be found in the epidemiological literature, as well. The lack of consistency among measured values contributes to the wide variability in results reported.



base of exposure can also be problematic. Some studies have used schools to assess exposure (Wjst et al. 1993) but these number fewer than those using the primary residence.<sup>2</sup>

Thus, questions still remain regarding the exact nature of the connections—are there health impacts? if so, which are the causal pollutants? which vehicles contribute most to health effects? how do traffic and meteorological parameters change observed concentrations? how far does roadway influence extend?<sup>3</sup> While the present work does not attempt to answer all of these questions, it involved creating and then assessing a large dataset to synthesize data from the literature on near roadway air pollution, and to determine what is known about the atmospheric transport and fate of air pollution, as emitted from vehicles on a roadway and carried downwind, eventually reaching background concentrations. We use existing data on the observed concentrations of particulate matter (PM) mass and number, criteria pollutants, and other potentially harmful traffic-related pollutants at or near roadways. Other studies have examined on-road (Westerdahl et al. 2005) and in-vehicle exposure (see, e.g., Chan et al. 1991; Zhu et al. 2007) and confirmed that such environments can constitute a large fraction of overall exposure to air pollution; this work extends understanding about pollutant exposure by pooling results across numerous studies and by using statistical methods to assess near road conditions. The results can be used to inform near-term policy.<sup>4</sup>

---

<sup>2</sup> The direct measurement of air pollutants outside and inside the home has also been undertaken. Fischer et al. (2000) found that homes classified as having high traffic exposure (mean 16,082 vehicles/day on nearest main street) exhibited higher concentrations of polycyclic aromatic hydrocarbons, volatile organic compounds, elemental carbon, and benzo(a)pyrene than those classified as low traffic (< 3000 veh/day on nearest main street). Differences ranged from 80-120 percent. Particulate mass was elevated to a lesser extent. Indoor contrasts varied similarly, supporting the use of some pollutants as indicators of traffic-related exposure, but not others.

<sup>3</sup> An additional concern relates to the *distribution* of adverse health effects. At present in California, near roadway exposures are borne disproportionately by citizens who are poor, minority, and of school age (Houston et al. 2004; Houston et al. 2006) raising important environmental justice concerns. Gunier et al. (2003) specifically investigated racial disparities in childhood exposure in California, determining that children of color were three times more likely than white children to live in high traffic areas. Seattle and Portland were also studied from this vantage point (Bae et al. 2007). The authors reached similar conclusions about the exposures of low-income communities of color to traffic-related pollution near roadways, finding that African American and/or poor populations disproportionately reside within 330 ft from roadways. A companion study to this investigation examines the community of Barrio Logan, in San Diego, California, where diesel truck traffic generated primarily by a local port facility creates a host of problems for residents (Karner et al. 2008).

<sup>4</sup> As of 2008, there were few federal or state policies that addressed air pollution and near road development; one guidance document was published by the California Air Resources Board and set a recommended minimum threshold for “siting new sensitive land uses within 500 feet of a freeway, urban roads with 100,000 vehicles/day, or rural roads with 50,000 vehicles/day (CARB 2005, p. 4).

U.S. federal air quality monitoring guidelines have generally focused on regional air pollution problems and have encouraged monitor placement away from the direct influence of major roadways. Collected monitoring data is typically used to characterize regional pollutant concentrations and to determine attainment status for the National Ambient Air Quality Standards (NAAQS). Lacking measured near roadway pollutant concentrations, most analysts have historically used dispersion modeling tools to examine near roadway concentrations (see, e.g, Holmes and Morawska 2006). This report advances understanding of pollution concentrations in the near road region by helping to characterize these concentrations using publicly available data. One of the benefits of presenting results from monitored data is that they can be used to validate modeling tools or to help assess under which conditions modeling results are most robust.

The remainder of this report is structured as follows: the next section presents an overview of the methods employed in the analysis. A summary of the literature follows. We then discuss how data was selected and prepared. We then discuss results, divided into three sections: actual (not normalized) concentrations, concentrations normalized to background, and concentrations normalized to the roadway edge. For both normalized sections, we perform two sets of analyses to examine the variation of observed concentrations with distance: a local regression which uses linear fits of subsets of the data to construct a smooth line, and an analysis of variance (ANOVA). A final results and discussion section compares both normalization techniques. The final two sections note limitations to the analysis and discuss policy-relevant conclusions.

We find that normalizing to the edge concentration performs better than normalizing to background concentrations in terms of making the data commensurate over time and over varying meteorological conditions. Nitric oxide (NO), total oxides of nitrogen (NO<sub>x</sub>), carbon monoxide (CO), elemental carbon, metal deposition, non-alkane hydrocarbons, and ultrafine particle numbers were all elevated in the near road region from 1.7 – 20 times above background values; most decayed to background concentrations by 150 m from the edge of road. Benzene and nitrogen dioxide (NO<sub>2</sub>) were elevated near roads (1.4 – 1.5 times above background) and reached background levels by 300 – 350 m, but the results were less conclusive for these pollutants. Other important criteria pollutants, such as particulate matter less than 10 µm in

aerodynamic diameter ( $PM_{10}$ ) and particulate matter less than  $2.5\ \mu m$  in aerodynamic diameter ( $PM_{2.5}$ ), consistently showed no or little relationship with distance from the road.

## 2 METHODS OVERVIEW

More than 30 studies from the peer-reviewed environmental and atmospheric science literature which contained data on observed near roadway concentrations were identified and assessed. The studies were drawn from a wider literature review which encompassed studies on near roadway issues in general—those which compared indoor and outdoor concentrations, conducted emissions measurements on a mobile platform (only used for this inquiry if the platform was stationary, not in moving traffic), measured in-vehicle concentrations, directly modeled near road concentrations, or measured concentrations in urban canyons. Beyond the peer-reviewed literature, several institutional reports were also evaluated. We made an extensive effort to identify and assess all relevant published materials; however, it is possible that additional studies might exist with near roadway measurements. Future work can incorporate new or previously undiscovered measurements to the extent possible.

The collected studies do not employ uniform sampling methods, nor do they provide data of uniform quality. A Microsoft Access database was therefore compiled to store relevant details of each study including details of the roadway, traffic volume and composition, the pollutants measured and averaging times, as well as observed concentrations. This analysis only employed the observed concentrations and their associated distances, but future work can use the full extent of collected data (see section 6 for a discussion). The open source statistical package R was used for subsequent data presentation and analysis. Code for the generation of figures and analysis results is included in Appendix A. Appendix B includes bibliographic material from the near road literature used in this study; Appendix C includes a description of the Access database created as part of this study.

Many factors affect observed near road concentrations; by aggregating near road data using two normalization procedures (described in detail in section 5.2.1), this study allows a broader understanding of near road trends and similarities. The present study is not meant to state that the concentration of a pollutant at a given distance from the road is a single value. Rather, the information assembled here provides a schematic guide for the near road zone. To our knowledge, this is the first such meta-analysis of near roadway air pollution data to

comprehensively portray a visual representation of the body of near roadway field work.<sup>5</sup> Two previous studies have attempted to summarize a similar body of work (Brugge et al. 2007; Zhou and Levy 2007).

Brugge et al. (2007) reviewed cardiopulmonary health risks associated with near road exposures and concluded, from a review of eight studies, that ultrafine particle number, black carbon, CO, and oxides of nitrogen (including NO and NO<sub>2</sub>) are elevated near roadways and that the most important zone of influence extends to those individuals residing 30 m from freeways compared to those living 150 – 250 m away.

Zhou and Levy (2007) performed a meta-analysis on the “spatial extent” of impacts resulting from mobile source air pollution. Spatial extent was defined as the distance at which the effects of the roadway was no longer observable in terms of pollution concentrations, or health effects, for example. They combined studies from five areas: observed pollutant concentrations, dispersion modeling, biomonitoring (including organic material uptake of pollutants near roads),

---

<sup>5</sup> Since the data were assembled from previously published studies, our methods can be described as a meta-analysis. One definition of a meta-analysis is offered by Glass (1976): ... [meta-analysis is] the statistical analysis of a large collection of analysis results from individual studies for the purpose of integrating the findings. It connotes a rigorous alternative to the casual, narrative discussions of research studies which typify our attempts to make sense of the rapidly expanding research literature. (p. 3)

Glass (1976) notes further that meta-analysis is best suited for areas of work which he describes as “outcome evaluation [rather] than analytic research” (p. 3). Analytic research is undertaken in disciplines which follow a cumulative, natural science model where the results of each study chart a more or less clear course for successive ones. Outcome evaluation refers to disciplines which operate on a model where research is completed simply to see what the outcome is. Applied disciplines such as public health or engineering often fit into this model. It is here where meta-analysis can bring together seemingly disparate results.

Previous meta-analytic work in the public health literature on air quality has focused on, for example, mortality due to particulate and gaseous air pollution. Schwartz (1994) collected 13 studies linking increased PM concentrations with risk of daily mortality. Pooling the results, which were conducted in a number of locations and meteorological conditions, allowed the author to conclude that the relationship was causal and not a result of confounders in the individual studies. Similar work by Stieb et al. (2002) pooled 109 studies that examined PM<sub>10</sub>, CO, NO<sub>x</sub>, sulfur dioxide (SO<sub>2</sub>), and ozone and their effects on mortality. The results showed significant increases in the risk of all-cause mortality with an increase in the 24 hour concentrations of each pollutant. The value of the increase was equal to the mean of a representative geographic sample determined by the authors based on their dataset. These two meta-analyses had clearly defined outcomes (mortality) as well as a causal mechanism to investigate (i.e., decreases in air quality cause increases in mortality). Another important example of this type of meta-analysis is the United States Environmental Protection Agency’s (EPA) rulemakings on the NAAQS; these involve vast literature reviews and meta-analyses to determine appropriate standards.

Other types of meta-analyses are possible which do not focus on a particular outcome measure, but rather seek to summarize existing data. Mudd and Diesendorf (2008) examined the sustainability of uranium mining and milling by examining reports from the uranium mining industry to summarize bottom-line sustainability metrics such as water use and carbon dioxide (CO<sub>2</sub>) emissions for various mining projects. They were interested in trends over time. While many factors affect these metrics, through collecting, plotting, and summarizing the data in tabular form, they concluded that energy and water consumption, and greenhouse gas emissions have been rising over time corresponding to a decrease in the quality of uranium ore.

land use regression modeling, and epidemiological assessments of elevated health risks. Their final dataset included information on the spatial extent obtained from 33 studies, in addition to characteristics about each study (e.g. study type, pollutant type grouped into four categories, the definition of spatial extent, meteorological data if available, and traffic information). Performing a univariate ANOVA with spatial extent as the dependent variable and each of the groups as independent variables in turn, they determined that three factors were important: the spatial extent definition, the pollutant type, and local meteorology. A limitation of the study was that the authors only used definitions of the spatial extent as determined in the individual studies, making it difficult to complete some cross-study comparisons. For example, one category of definition for monitoring studies was the percentage change from a high value near the road to some percentage downwind. This percentage varies across studies, and the spatial extent varies accordingly.

Zhou and Levy (2007) also completed a dispersion modeling case study to accompany their assessment of monitored concentrations. Their modeling results show that, in order to decrease the percentage of the near road pollutant concentration high from 50 to 10 percent, distance from road (i.e., spatial extent) needs to increase from 60 to 380 m for non-reactive pollutants.

Zhou and Levy (2007) extracted the spatial extent from each of the studies they examined, but did not normalize comparisons across studies. As such, their analysis is highly aggregated. Our study complements the findings prepared by Zhou and Levy (2007); our work focuses on examining literature on observed concentrations, and records individual observations of the concentration to allow a normalized assessment between studies. We are also able to overcome another shortcoming noted by Zhou and Levy (2007), namely that data limitations prevented them from disaggregating results by particulate size and mass fractions. Our dataset includes enough measurements that meaningful disaggregation of particle size and mass is possible.

Appendix B presents an annotated bibliography of the near road literature included in our assessment; it focuses on studies reporting field measurements of near road concentrations. It summarizes findings, noting which pollutants were studied, how fast pollutants decayed to background concentrations, and other items of note. Appendix B also includes notes regarding

unique issues from studies that affected their treatment when assembling the database described in Appendix C.

### 3 LITERATURE SUMMARY

The near road studies reviewed represent wide geographic, meteorological, and traffic operational variation. This summary narrows the discussion to only those studies which were recorded in our database.<sup>6</sup> Section 4 details how data were aggregated and prepared for subsequent analysis. Our analysis unit was one distance/concentration pair. As shown in Table 1, 572 such pairs were identified from 34 papers spanning 95 unique *measurement sets*. A measurement set is a group of distance/concentration pairs originating from the same study and representing the same pollutant and different measurement conditions. Many studies reported results from different observation days, seasons, or traffic conditions. If these data were not combined in the reported results of the study, they were recorded as separate measurement sets for analysis.

**Table 1: Summary of Observations**

	Studies	Number of	
		Separate measurement sets <sup>a</sup>	Distance/Concentration Pairs
Background		75	380
Trend		57	480
Total	34	95	572

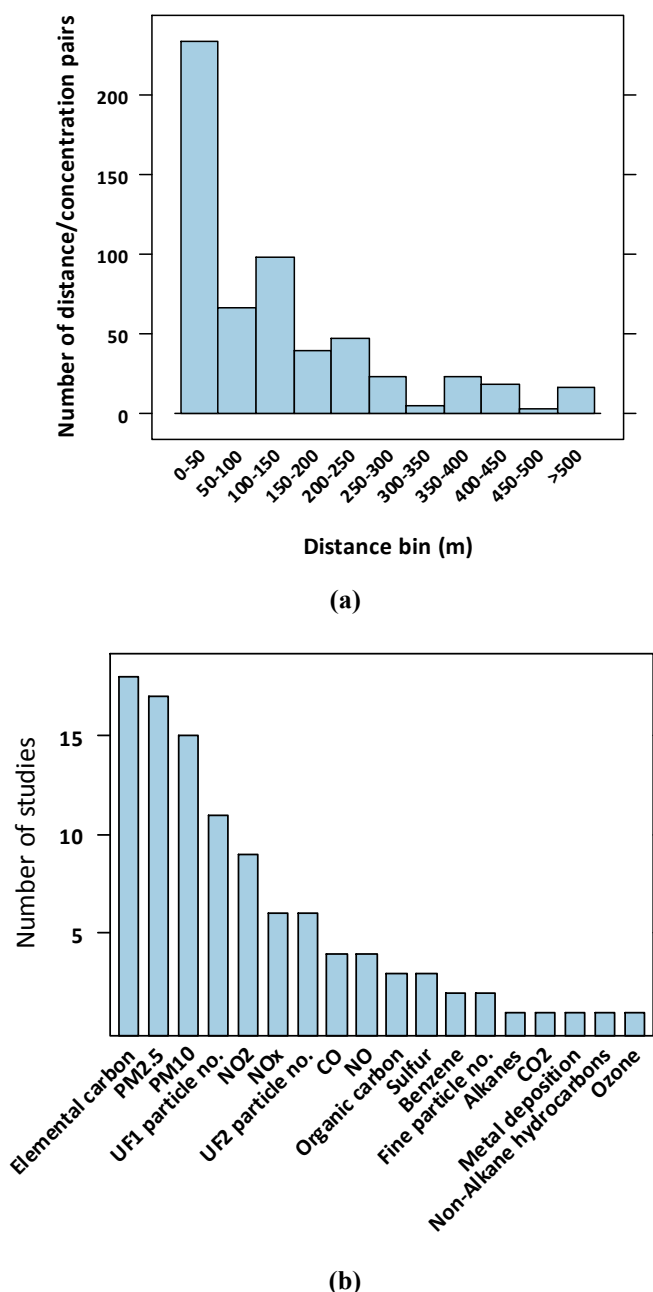
**Note:** Since many studies perform multiple observations, not all of them commensurate, it is not meaningful to divide them into those that measured background concentrations or presented trend measurements, hence the shaded grey area.

<sup>a</sup>A measurement set is defined here as a group of distance/concentration pairs originating from the same study and representing the same pollutant but different measurement conditions; sets were typically counted as a unique series of measurements when they were reported as taken at different times or under different conditions. Many studies report more than one separate measurement set.

As indicated by Figure 1a, the overwhelming majority of observations were conducted within 150 m of the roadway. This reflects the common view of near roadway pollution, as codified in some land use planning documents. For example, the California Air Resources Board (CARB) recommends a 500 foot (150 m) setback for sensitive land uses (CARB 2005). When grouped by pollutant types included in each study, Figure 1b shows an emphasis on PM mass (30 percent of all studies), oxides of nitrogen (18 percent of all studies), particle number

<sup>6</sup> More than 75 papers were initially included for review, but less than half had data suitable for entry into the database. These were instead used as supporting evidence for the literature review in Appendix B.

concentrations (18 percent of all studies), and elemental carbon and surrogates (17 percent of all studies).



**Figure 1: Summarizing the data: (a) histogram of all observations grouped by 50 m distance bin (b) barchart showing number of studies grouped by pollutant type.**

**Note: “Elemental carbon” in (b) refers to both direct and surrogate measures of that pollutant, and measures of particle number (UF1, UF2, fine) are discussed in section Error! Reference source not found.. The sum of the bars in (b) total in excess of the 34 papers included in our database since many papers studied more than one pollutant and are thus counted more than once in that figure.**

Particulates dominate our database, with mass and number concentrations observed in nearly half of all studies. This tracks a source of uncertainty in the epidemiological literature—namely that the source of particulate-related health risks is still under investigation. Possible contributing factors are total number concentrations, sizes, surface area, and composition (Nel 2005)—with number and surface area likely playing a key causal role (Donaldson et al. 2001). As of the late 2000s the EPA did not regulate PM number concentrations, and instead focused on PM<sub>10</sub> and PM<sub>2.5</sub> mass concentrations. The existing mass-based PM regulatory approach may prove to be problematic if epidemiological work establishes a definitive link among ultrafine particle number concentrations, exposure, and adverse health impacts. As shown below, particle number concentrations are often substantially elevated in the areas near roadways, while PM mass is a regional-scale pollutant that is not generally higher, or is only slightly elevated, near roads.

## 4 DATA SELECTION AND PREPARATION

Field studies to measure near road concentrations have used diverse measurement approaches and technology; typically, however, the studies arrange emissions monitoring equipment along an approximately perpendicular vector emanating from the facility. Studied distances and pollutants varied as well, and data was sometimes collected as part of a model verification or calibration exercise. If a paper described the collection of measurements at different locations, or under different conditions, we categorized the results as constituting separate measurement sets and included them as such in the database. The collected measurements typically represent sustained measurement campaigns over periods extending several weeks to one year.

A common result in the literature highlights the importance of the meteorological parameters wind speed and direction, for pollutants with near road variation.<sup>7</sup> Intuitively, with wind from the road to the receptors, concentration gradients are more pronounced and extend to a greater distance than with wind parallel to or away from the receptors (Hitchins et al. 2000).

---

<sup>7</sup> Other parameters are also important. As noted by Morawska et al. (1999) in their study of near roadway particulates in the Netherlands, “in close proximity to the freeway particle concentrations are elevated, but after a certain distance from the freeway there is no clear relationship between distance and particle concentration. This distance may not always be the same but rather can be related to the topography of the location, the meteorological conditions and the intensity and presence of other particle sources” (p. 1268).



All else equal, pollutant concentrations at downwind receptors reach a maximum at some wind speed, then decrease as wind speed either increases or decreases from this maximum value. In general, at some distance downwind, observed concentrations decay to the background value, although under some conditions, higher concentrations persist beyond 500 m, specifically at night (Zhu et al. 2006). Data was only entered into the database for concentrations measured when wind was approximately from the road<sup>8</sup> or was aggregated over a variety of meteorological conditions.

The reported data from the studies are, in many cases, limited by the absence of basic information, compounding the difficulty of their inclusion in the database. Similar difficulty with this literature was noted by Zhou and Levy (2007). Such information includes, but is not limited to, a fixed distance reference point, disaggregated concentration measurements, or measurement averaging times (i.e., the time periods represented by the measurements: peak period, all day, several days, and so on). Additionally, the number of distances at which measurements were taken varied widely. For example, many studies measured only one data point close to the roadway edge (typically within 10 m) with the goal of correlating variations in traffic and meteorological parameters with concentrations observed at that monitoring location (Ruellan and Cachier 2001; Molnar et al. 2002). These were generally not entered into our database, since they are not amenable to normalization. When choosing which data to include, all data was entered except for that which explicitly stated it was from specific wind regimes that would obscure concentration gradients. It was not possible to, for example, only include studies which monitored during the peak period to account for the influence of traffic on observed concentrations, since not all measurement times were known for all studies. Even though some variability will be less pronounced as a result of including different conditions in the same dataset, on net, we find that the normalization procedures (described below) adequately addressed variation among data collection protocols.

Our analysis takes several approaches to mitigate possible disparities in the pooled data. It was important to ensure that all data be presented with respect to the same reference point. In cases where the distance reference point (either center line or edge of road) was not specified, the edge was assumed, since it was the more common. All observations are presented assuming that

---

<sup>8</sup> This does not refer to only perpendicular winds. Maximum concentrations are reached when wind has an angular component, since the effective area of the road from which material can be transported increases from perpendicular to a maximum before decreasing as wind approaches parallel.

the zero distance refers to the edge of the traveled lanes. It was sometimes not possible to ascertain the existence of a shoulder; however, since a typical maximum shoulder width is 3 m (Caltrans 2008, Table 302.1), the error should be small. If the roadway cross-section was not provided, an estimate of the distance from center to edge was determined using Google Maps to measure distance based on the best description of the monitoring location given in the study. Units of measure were also standardized across each pollutant type, although this is only relevant for presentation of the unnormalized data.

We performed several grouping operations to collapse the data. These enhanced our ability to complete statistical tests, and enabled observations by individual pollutants. The following describes our treatment of data by pollutant.

Elemental carbon can be measured by many different means. Cyrus et al. (2003) notes the importance of this pollutant as a surrogate for diesel particulate matter (DPM). They noted further that the terms black carbon and soot are used interchangeably with elemental carbon, but that the two pollutants have somewhat different behavior. Direct measurement of elemental carbon is complex, but the reflectance of particulate filters is correlated with elemental carbon, regardless of the upper size cut of the filter used (whether reflectance is measured from PM<sub>10</sub> or PM<sub>2.5</sub> filters, the two recorded in our database) since elemental carbon resides mostly in the submicron range (Fischer et al. 2000). Measurements of black smoke also are correlated with elemental carbon. We combine all of these indicators together, labeled as “elemental carbon” in figures but frequently referred to as “elemental carbon and surrogates” in the text.

Total suspended particles (TSP) are converted to PM<sub>10</sub> using a factor of 0.55 to estimate PM<sub>10</sub> concentrations following Levy et al. (2000).

Alkanes (ethane, propane, *n*-butane, and *iso*-butane) and non-Alkane hydrocarbons (ethane, acetylene, *iso*-pentane, benzene, propene, *n*-pentane, and 1,3-butadiene) were studied by O’Donoghue and Broderick (2007) who noted that those groupings tended to behave similarly—alkanes did not vary substantially with distance, while the non-alkanes did. Given the regulatory interest in benzene as a mobile source air toxic (MSAT), we extract benzene from the non-alkanes to place it into its own group, but keep the others grouped together since they were the only measurements taken of those pollutants. If future studies produce more of these measures, future work can split data on these pollutants into separate groups.

The term “ultrafine” typically refers to particles less than 100 nm in diameter (Donaldson et al. 2001), and particle number concentrations (as opposed to mass concentrations) are typically used to quantify ultrafine pollution. Detectors used for counting particle number varied among scanning mobility particle sizers (SMPS) fitted with condensation particle counters (CPC), stand-alone CPCs, condensation nucleus counters, and optical particle counters. The size range measured by each can be set by the analyst, and since the number concentration tends to increase depending on the low end of the measured range (Molnar et al. 2002),<sup>9</sup> we grouped studies into three categories according to that value.

The size range measured for each study that analyzed particle number concentrations is shown in Table 2; many studies measured multiple size ranges. Although some studies set an upper size limit in excess of 100 nm when defining ultrafine particles, most studies report ultrafine particle number concentrations for particles < 100 nm in aerodynamic diameter. During our assessment, we defined three groups of monitored particulate data in addition to PM<sub>10</sub> and PM<sub>2.5</sub> data: we labeled data “UF1 particle number” when the data began at 3 nm, “UF2 particle number” when the data began at 15 nm, and “Fine particle number” for particulate data beginning in the 300 nm size range (~ 0.3 µm, just above the ultrafine classification). While the SMPS can measure size distributions as well, these data are not suitable for our database, since the output is not a single measurement at a distance from the road, but rather a distribution.

---

<sup>9</sup> For example, Shi et al. (1999) found comparatively little variation for particle number concentrations in the size range greater than 72 nm when compared to that greater than 30 nm, and Nanzetta and Holmen (2004) found higher concentrations in the range from 6 – 237 nm than 23 – 898 nm.

**Table 2: Particle Size Ranges Counted in Studies Which Recorded Particle Numbers**

<b>Particle Number Classification</b>	<b>Size Range Measured</b>	<b>Study</b>
UF1	> 3 nm	Gidhagen et al. (2004)
UF1	> 6 nm	Ntziachristos et al. (2007b)
UF1	6 – 220 nm	Zhu et al. (2002a); Zhu et al. (2002b)
UF1	6 – 237 nm	Nanzetta and Holmen (2004)
UF1	4 – 163 nm	Gramotnev and Ristovski (2004)
UF1	6 – 300 nm	Zhu et al. (2008)
UF1	> 7 nm	Weijers et al. (2004)
UF1	7.4 – 290 nm	Zhu et al. (2006)
UF1	> 8.2 nm	Kuhn et al. (2005a); Kuhn et al. (2005b)
UF1	9.6 – 352 nm	Shi et al. (1999)
UF2	15 – 697 nm	Hitchins et al. (2000)
UF2	16 – 638 nm	Ntziachristos et al. (2007b)
UF2	20 – 1000 nm	Reponen et al. (2003)
UF2	23 – 898 nm	Nanzetta and Holmen (2004)
UF2	20 – 1000 nm	Hagler et al. (2008)
Fine	500 – 20,000 nm	Hitchins et al. (2000)
Fine	300 – 20,000 nm	Reponen et al. (2003)

## 5 RESULTS AND DISCUSSION

The results and discussion focus on the visualization of the data, coupled with statistical analysis. Results are shown for the unnormalized concentrations, as well as those normalized to background concentrations and edge of road concentrations. Statistical and normalization methods are also elaborated upon in this section.

### 5.1 Unnormalized near roadway concentrations

Figure 2 shows all of the collected data separated into panels by each of the 21 available pollutant types. All elemental carbon surrogates are plotted separately. This figure is not normalized, and each panel has unique axes with units that vary according to the pollutant type. Unique plotting characters represent a single measurement campaign conducted within one study. Concentration measurements for the same pollutants taken at different facilities have wide variation and some pollutants have more data than others (also illustrated in Figure 1). There is substantial variation in concentrations as distance from roadway decreases. For example, measurements of elemental carbon adjacent to roads range from  $7.34 \mu\text{g}/\text{m}^3$  observed by Lena et al. (2002) for the Hunts Point community in the Bronx, New York to the low of  $0.7 \mu\text{g}/\text{m}^3$

observed by Ntziachristos et al. (2007b) on a rainy day next to the I-710 freeway in Los Angeles. Both areas are heavily influenced by diesel trucks, for which elemental carbon is a known marker, but the meteorology and local traffic conditions clearly affect the measured values. Additionally, the studies had different objectives—Hunts Point was studied in response to community concerns and to quantify relationships to heavy truck activity, while I-710 was studied in this case to classify the physical and chemical properties of particles near the freeway in detail. In the former case, measurement of a background concentration was essential to quantify the community's incremental exposure above background, while in the latter case it was not. Differing reasons for near road data collection can determine the study methods and data collected.

Similar disparities in the range of observed concentrations appear throughout the data. Table 3 contains the sample size, range and mean of observed concentrations, and coefficient of variation grouped over the entire distance range and calculated individually for each pollutant. This table does not include background concentration measurements. Of course, the range of observations is affected by the distance over which observations were made. One of the metrics we examined in particular was the coefficient of variation,<sup>10</sup> since it provides a measure of the spread of observations that is not dependent on units and can therefore be compared across pollutants. A reasonable expectation might be that, as the number of observations increases, the coefficient of variation increases as well since observed concentrations will differ between study locations and at different distances. However as shown in Table 3, there is no visible trend with coefficient of variation as the number of data points increases. Indeed, all of the concentration measurements are highly variable. The coefficient of variation might also be affected by the extent to which a given pollutant varies with distance. Those pollutants which vary little may exhibit a lower coefficient. Sorting the table in order of increasing coefficient of variation does not yield any new insights, however. Others have observed that the coefficient of variation undertaken at a single measurement site can be quite stable, even under changing seasons (Vardoulakis et al. 2005).

---

<sup>10</sup> The coefficient of variation is the standard deviation divided by the mean.

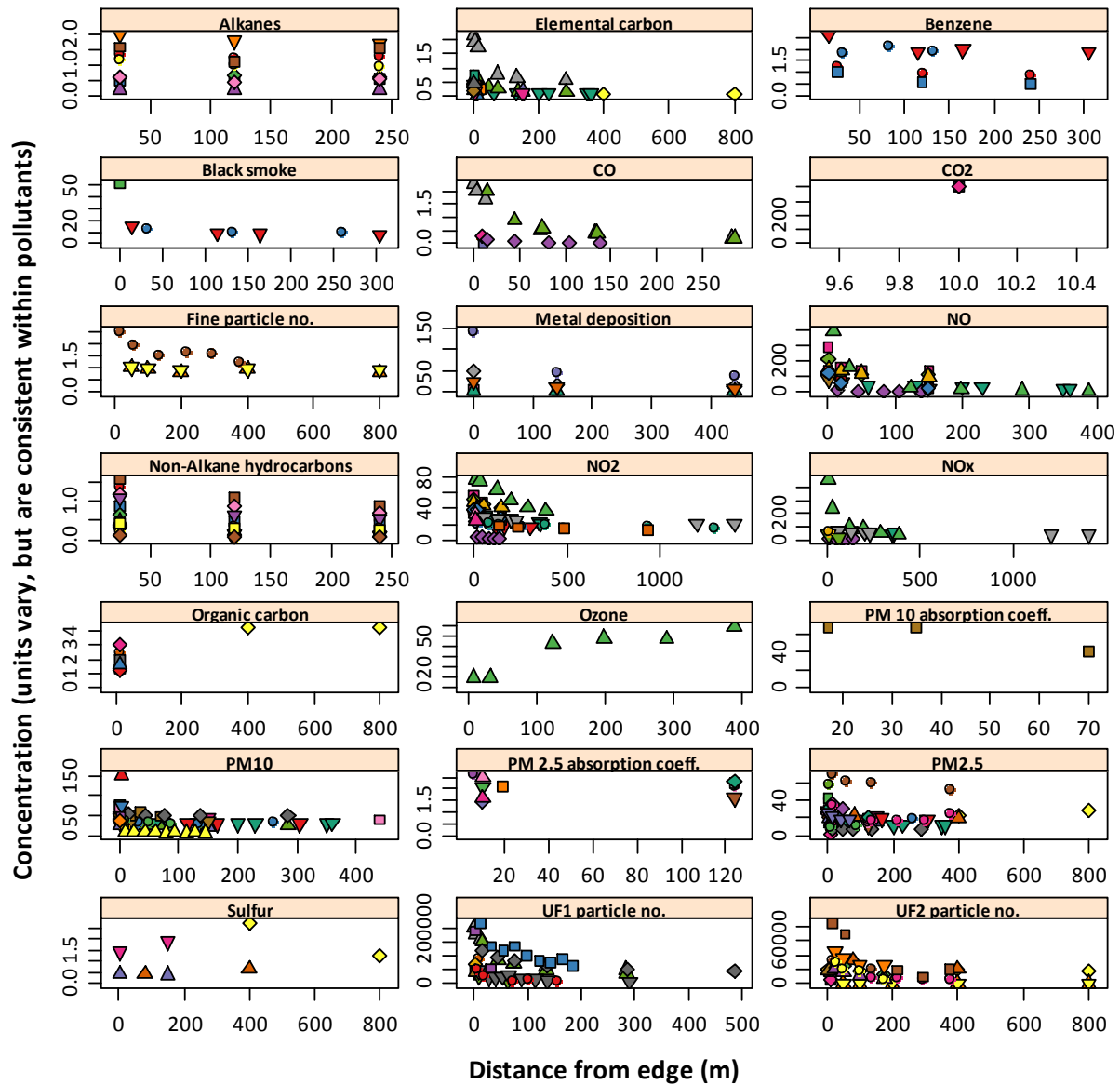


Figure 2: All collected near road air pollution data plotted in separate panels.

Note: These data are not normalized. Each panel has unique axes to highlight the maximum possible range of data. Within each panel, different plotting characters represent unique datasets.

Units are as follows: Alkanes (ppb), Elemental carbon ( $\mu\text{g}/\text{m}^3$ ), Benzene ( $\mu\text{g}/\text{m}^3$ ), Black smoke ( $\mu\text{g}/\text{m}^3$ ), CO (ppm), CO<sub>2</sub> (ppm), Fine particle no. ( $\text{cm}^{-3}$ ), Metal deposition ( $\mu\text{g}/\text{m}^2/\text{d}$ ), NO (ppb), Non-Alkane Hydrocarbons (ppb), NO<sub>2</sub> (ppb), NO<sub>x</sub> (ppb), Organic carbon ( $\mu\text{g}/\text{m}^3$ ), Ozone (ppb), PM<sub>10</sub> absorption coeff. ( $10^{-6}/\text{m}$ ), PM<sub>10</sub> ( $\mu\text{g}/\text{m}^3$ ), PM<sub>2.5</sub> absorption coeff. ( $10^{-5}/\text{m}$ ), PM<sub>2.5</sub> ( $\mu\text{g}/\text{m}^3$ ), Sulfur ( $\mu\text{g}/\text{m}^3$ ), UF1 particle no. ( $\text{cm}^{-3}$ ), UF2 particle no. ( $\text{cm}^{-3}$ ).

**Table 3: Summary Statistics Calculated Over the Entire Distance Range Grouped by Pollutant Type**

<b>Pollutant (units)</b>	<b>n</b>	<b>Min</b>	<b>Max</b>	<b>Mean</b>	<b>Coeff. of Variation (%)</b>
CO <sub>2</sub> (ppm)	3	402	426	411	3.2
PM10 absorption coeff. (µg/ m <sup>3</sup> )	3	40.8	67.6	58.3	26
Ozone (ppb)	6	10	59	36.2	58
Sulfur (µg/ m <sup>3</sup> )	8	0.39	2.73	1.16	73
Organic carbon (µg/ m <sup>3</sup> )	8	1.2	4.3	2.51	50
Black smoke (µg/ m <sup>3</sup> )	8	7.4	51	15.03	98
PM <sub>2.5</sub> absorption coeff. (µg/ m <sup>3</sup> )	10	1.42	2.6	2.07	19
Benzene (µg/ m <sup>3</sup> )	13	0.5	2.6	1.49	44
Metal deposition (µg/m <sup>2</sup> /d)	15	1	144	24.3	152
Fine particle no. (/cm <sup>3</sup> )	16	0.8	2.5	1.22	39
CO (ppm)	19	0.004	2.3	0.63	122
Alkanes (ppb)	24	0.2	1.99	0.93	58
NO <sub>x</sub> (ppb)	26	1.5	460	66.9	138
Non-Alkane hydrocarbons (ppb)	36	0.05	1.59	0.50	80.5
NO (ppb)	41	0.16	390	80.8	103
Elemental/Black carbon (µg/ m <sup>3</sup> )	43	0.34	21.7	3.93	120
UF2 particle no. (/cm <sup>3</sup> )	50	0.08	85000	16800	98
UF1 particle no. (/cm <sup>3</sup> )	51	7500	218000	69500	77
PM <sub>10</sub> (10 <sup>-6</sup> /m)	58	7.3	150.8	35.6	63
PM <sub>2.5</sub> (10 <sup>-5</sup> /m)	66	2.1	69	21.6	64
NO <sub>2</sub> (ppb)	68	1.2	78	29.2	55

**Note:** Minima, maxima, mean, and the coefficient of variation are calculated from all concentration values from that pollutant. The number of data points, n, is given by the total number of distance/concentration pairs used in the calculation.

## **5.2 Normalized near roadway concentrations**

### **5.2.1 Normalization methods**

Because of the high variation in observed concentrations both between and within pollutants, the analyses undertaken below used two normalization procedures to facilitate meaningful comparisons among the data: normalization to background and normalization to edge of road concentration. A synthesis of the results from both methods is presented in section 5.2.4.<sup>11</sup> If all information contained in the database were collected under similar conditions and

<sup>11</sup> Two separate datasets were created from the raw data to contain all data normalized using each method. Each of these separate datasets contained an aggregation of individual sets of measurements from each study included in the database. The normalizations were only performed on one set of measurements at a time. Some overlap was possible

using similar protocols, only one type of normalization would be required to compare studies on an equal footing. Perfect information about the decay profile would constitute an edge concentration, plus several measurements at a perpendicular transect, with the final measurement representing a background value. From such (hypothetical) data, the near road increment above background would be known, as would the percentage decreases with distance and the distance at which background concentrations are reached. However, many studies are missing information on background concentrations, for example, or only took concentration measurements at two points, making an exponential fit impossible. Employing both normalization methods allows all of the data to be used, one way or another, to determine which pollutants concentrations are higher near roads, the range over which those increased concentrations occurs, and the rate at which concentrations decay to background values.

Data is sometimes normalized to a unit such as wind speed or traffic volume (e.g., Zhu et al. 2008) or by subtracting the background concentration (e.g., Weijers et al. 2004; O'Donoghue and Broderick 2007). However, it was not possible in this study to normalize to traffic volume or specific meteorological conditions, since those normalizations required access to data often not documented in the published literature.

### **5.2.1.1 Overview Discussion: Normalized to Background**

Our first normalization technique—normalized to background—performs a similar operation to background subtraction but puts all of the pollutants on nondimensional terms. It utilizes studies that provided background (upwind) measurements. For these studies, the background concentration was used as the basis for normalization—all other concentrations observed for each set of measurements were divided by the background value.

The background normalization procedure directly yields the *relative* concentrations of pollution in the near roadway zone as compared with areas where traditional monitors would be sited—outside the areas of influence of point sources of pollution such as power plants, or line sources such as roads. Investigating trends using this normalization procedure will show where concentrations (and thus health risks potentially caused by near roadway exposures to those concentrations) would drop to the background level. Very high or low background concentrations in specific locations could obscure or overstate trends with this normalization

---

between the two datasets, if individual sets of measurements contained the information necessary to facilitate both types of normalization, but in general each dataset was treated separately.



technique.<sup>12</sup> High background concentrations will tend to generate flatter gradients and low background concentrations will generate steeper gradients. These issues arise and are discussed below.

Since background normalization does not mitigate variability between studies, there is a danger that such variability will be so high as to make comparisons between the same pollutants calculated using background normalization meaningless. However, this should be apparent from simply looking at the data. To illustrate an example of how this would occur, the differences between the two procedures are discussed for the case of NO. Two sets of measurements with background concentrations are included in the pooled data for this pollutant. One was completed in the late 1970s (Rodes and Holland 1981) and the other in 2004 (Kim et al. 2004). The entire set of measurements from the 1970s is quite high, ranging from 290 ppb at 8 m from the road, to 5 ppb at 388 m with a 1 ppb background value. Dividing by the edge concentration brings all observations down. Even though the 2004 study shows a much lower roadway increment (17 ppb background with 38 ppb at 60 m from the road), when normalized, the shape of both curves is similar. But, dividing by background concentrations makes them very different—in the 1970s study, concentrations were about 150 – 200 times greater than background at 8 m from the road, whereas in 2004, they are only approximately twice background. The normalization to edge thus appears to create a more common comparison, since it leaves only variation by distance. As shown in Appendix B, most of the measurements are from the late 1990s through the present, so such stark differences are rare, although they do occur and are discussed in the text.

#### **5.2.1.2 Overview Discussion: Normalized to Edge**

Our second normalization technique—normalized to edge of road—uses the existing data to estimate an edge of road concentration according to the following algorithm. Assuming a 3 m shoulder width, if a measurement point was present in the first 3 m from the road, it was taken to represent an edge-of-road concentration. Otherwise, if there were only two data points in the set they were simply connected and their intercept with the vertical axis was used as the edge concentration. If the analysis of background concentrations and knowledge from the literature review agreed that, for a particular pollutant, there was no relationship between concentration

---

<sup>12</sup> O'Donoghue and Broderick (2007), for example, surmise that the flat concentration gradient for benzene observed by Roorda-Knappe et al. (1998) was due to a high background concentration in the Netherlands. In future cases where the existence of trends might be called into question, background concentrations should be measured and trends (or lack thereof) should be assessed with reference to the academic literature, and studies such as this one.

and distance from road, linear regression was used on the entire set of measurements from that series to determine the edge concentration. If three or more near road points were given and the background normalization and literature agreed that the pollutant varies with distance, an exponential decay model was utilized which varied according to distance. Previous work has shown that an exponential decay fits the atmospheric fate of pollutants which vary by distance (Zhu et al. 2002a; Zhu et al. 2002b; Gramotnev and Ristovski 2004). We fitted an exponential decay model to each set of measurements that met our criteria; the general model took the form shown in Equation 1,

$$C = \beta_0 + \beta_1 e^{-D/\theta} + \varepsilon \quad (1)$$

where  $C$  is the concentration value predicted based on  $D$ , distance,  $\varepsilon$  is a normally and independently distributed error term,<sup>13</sup> and  $\beta_0$ ,  $\beta_1$ , and  $\theta$  are estimated model parameters whose initial values were estimated using an algorithm described by Venables and Ripley (2002, p.216).<sup>14</sup> Representing concentration as a function of distance, the algorithm proceeds as follows. A quadratic is fit to distance, and three values for concentration,  $C_0$ ,  $C_1$ , and  $C_2$  are determined at equally spaced distance points:  $D_0$ ,  $D_1 = D_0 + \delta$ ,  $D_2 = D_1 + 2\delta$ . Substitution of these points

into Equation 1 yields an initial value of  $\theta$ , i.e.  $\theta_0 = \frac{\delta}{\ln\left(\frac{C_0 - C_1}{C_1 - C_2}\right)}$ . Ordinary least squares

regression of concentration on  $e^{-D/\theta}$  can then be used to determine initial estimates of  $\beta_0$  and  $\beta_1$ . Letting the parameters vary by modeled dataset recognizes that while distance is the most important independent variable<sup>15</sup> the precise shape of the curve will vary according to site- and study-specific details.

---

<sup>13</sup> It is likely that the error term is in fact heteroskedastic (meaning that its variance changes depending on the location of the observation), since its variance should increase as distance decreases, corresponding to a higher influence of the variation associated with roadway activity and as the magnitude of observed concentrations increases. Since the model is being used to describe a previously identified physical relationship, and not in an exploratory manner, and since we are not assessing model performance, this should not be an issue. Nevertheless, additional studies utilizing this type of normalization should examine this issue in greater detail, possibly applying a generalized model which allows the variance to be defined as a function of the predicted variable.

<sup>14</sup> Implemented in R as “negexp.SSival” in package MASS.

<sup>15</sup> Confirmed by previous results which modeled land use and traffic emissions as well as atmospheric dispersion models (see, e.g., Levy et al. 2003; Sapkota and Buckley 2003; Cohen et al. 2005; Gilbert et al. 2005; Kwon 2005; Luginaah et al. 2006; Gilbert et al. 2007; Ross et al. 2007).

Numerical methods are required for determination of the final coefficients for the nonlinear regression since the equations for their prediction are typically themselves nonlinear. The open source statistical package R includes three algorithms for nonlinear model estimation under the function `nls()`: Gauss-Newton, Golub-Pereyra, and “NL2SOL” from the package “Port.” An applied approach was taken to selection. Not all algorithms could be applied to each set of measurements—some failed to converge, or iterated indefinitely without finding acceptable parameters. Trial and error revealed that Golub-Pereyra functioned adequately most often and thus was used in the broad majority of cases while NL2SOL was able to estimate the remaining models. Two datasets from the series of non-Alkane measurements were not amenable to estimation regardless of algorithm, but showed relationships with distance from road upon visual examination of the data, with the nearest to road concentration high and the other two measured concentrations effectively equal. Three data points were available for each of these two remaining datasets, and the two closest to the road were fit linearly which provided a reasonable approximation for the edge of road concentrations. In all other cases, modifying the solution algorithm allowed the determination of a regression equation, even considering the small sample sizes ( $n = 3$ ) in some cases. For each fitted model, the prediction at zero distance—the edge of road concentration—was assigned a value of unity, and the subsequent measurements were normalized to it. These two normalization procedures—background and edge—allow all of the data to be assessed, regardless of other shortcomings.

Normalizing to the roadway edge should tend to group the data more closely together since we would expect that wide variations in magnitude would be reduced to a larger extent than in the background normalization case. From Figure 1, Table 1 and previous work, we know that differences in meteorology (Kuhn et al. 2005b; Ntziachristos et al. 2007b), traffic volume and fleet composition (Fischer et al. 2000; Janssen et al. 2001; Sapkota and Buckley 2003), and other factors (Baldauf et al. 2008b) can contribute to differences in observed concentrations of traffic-related air pollution. However, to the extent that we expect the *shape* of the decay curve to be roughly similar within each pollutant type we would expect dividing by the roadway edge concentration to preserve the shape while removing the magnitude since all observations will be compared on a common scale equal to unity at the edge of road. This intuition has been confirmed by recent work on the influence of roadway configuration and sound/vegetation barriers on observed concentrations (Baldauf et al. 2008b) and in other normalized comparisons

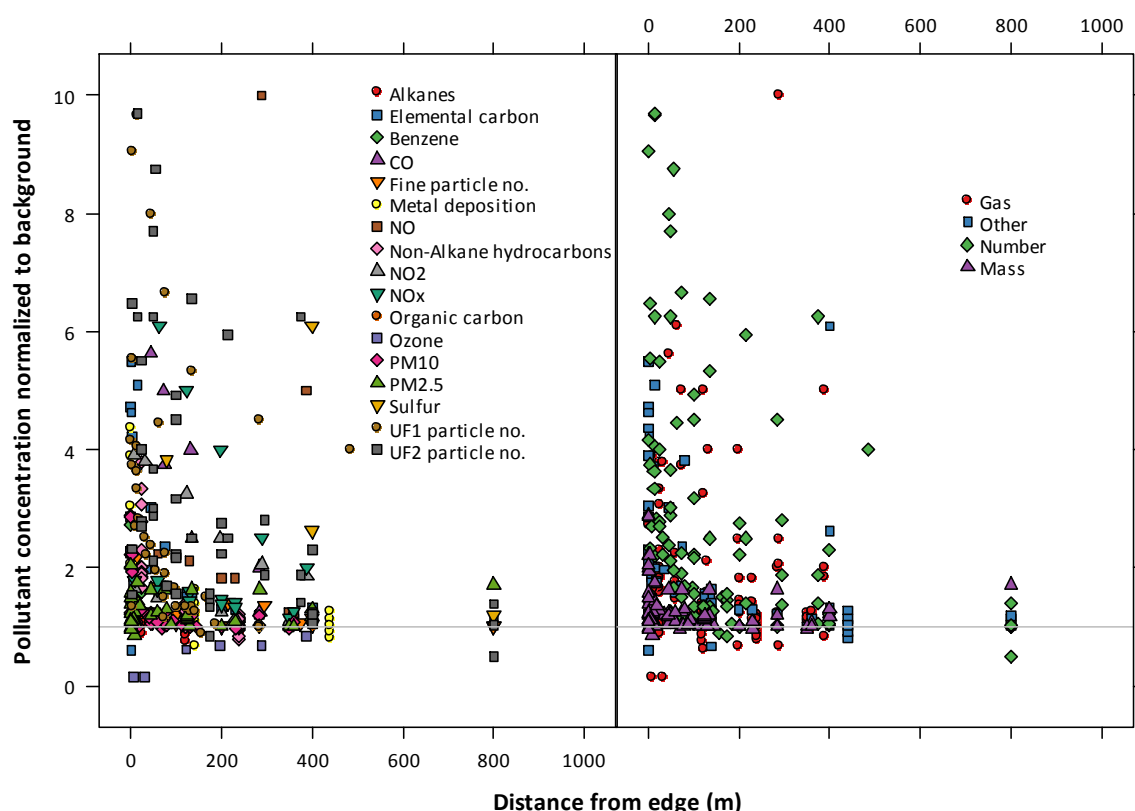
undertaken by Zhu et al. (2008) for three facilities normalized to unit wind speed and traffic volume. The result directly available from this normalization procedure is the *percentage decrease* in pollution concentration as measured from the roadway edge to the distance of interest. However, the distance at which background concentrations are reached can be gleaned from this normalization as well. If edge normalized concentrations are plotted against distance, the leveling of the curve will indicate that the roadway's influence has ceased and that background concentrations have been reached.

### 5.2.2 Results: Near roadway concentrations normalized to background values

Normalization of observed concentrations to background levels facilitates direct analysis of the distance at which pollutants reach background. While this does not allow us to infer that human exposure has also dropped to a background level at this distance since exposure is determined by many factors such as time of day and location of activities (Sexton and Ryan 1988), it can help for planning buffer zones to avoid exposure to the highest pollution levels. Figure 3 presents the 380 distance/concentration pairs with associated background concentrations normalized by the background value and grouped both by pollutant type (in the left panel) and further by pollutant “class” (right panel). Class comprises four categories: gaseous concentrations (denoted “Gas” and typically given in “parts-per” notation), particle number concentrations (denoted “Number” and typically given as tens of thousands of particles per  $\text{cm}^3$ ), particle mass concentrations (denoted “Mass” and typically given as  $\mu\text{g}/\text{m}^3$ ), and a final category for the remaining pollutants (denoted “Other” and typically given as  $\mu\text{g}/\text{m}^3$ ) which included elemental carbon, metal deposition, organic carbon, and sulfur). Wide variation is evident among the groupings. This variation is exacerbated in the case of background normalization, since differences due to the magnitude of the observation are not eliminated by normalizing, as previously discussed.

In general, the range of measurements widens closer to the road, although this might be an artifact of having more measurements in this zone. Ultrafine particle numbers, oxides of nitrogen, and CO appear to be most elevated adjacent to the road, reaching up to 10 times background concentrations. These three pollutant categories are clearly linked to traffic-related pollution. The highest density of points is visible between 1 – 4 times background concentrations as far as 250 m from the edge of road which potentially indicates the scale of high exposure. Above four times above background concentration appears to be an abnormally high

concentration. Elemental carbon,  $\text{NO}_x$ , and ultrafine particles mostly populate this range. When grouped by class, further patterns are evident. Mass concentrations are grouped closely around unity, although there is a high density of mass concentrations elevated up to twice background concentrations immediately adjacent to the roadway edge. Grouping together all measures of particle number (identified as “Number” in the right panel of Figure 3) and gaseous concentrations (identified as “Gas” in the right panel of Figure 3) shows no clear trends with distance. The analysis below will further disaggregate the pollutants to observe trends.



**Figure 3: Background normalized near road air pollution data.**

**Note:** Background concentration (a value of unity on the ordinate) is indicated by the solid gray line. The left panel shows all pollutants by pollutant type, while the right panel collapses pollutants into their “class” as previously defined.

### 5.2.2.1 Trend analysis

Figure 3 points to the possible existence of trends within the data, but is so highly aggregated that it is difficult to extract specific details from it. Statistical significance of trends is rarely assessed in the near road literature. This section makes a contribution in this regard, and investigates individual trends within each pollutant in two ways. The first uses a technique

known as locally weighted regression, or loess,<sup>16</sup> which reveals trends in the data. The second is an ANOVA. We provide an overview of each method and note important caveats that might affect the interpretation of results. The methods discussed in this section are identical to those employed in the edge normalized case, discussed beginning in section 5.2.3.1.

#### 5.2.2.1.1 *Local regression*

The primary benefit of local regression (loess) is that it does not presuppose a functional form of the relationship between dependent and independent variables (Cleveland 1979; Cleveland and Devlin 1988). This makes it capable of estimating an infinite number of smooth functions with almost complete freedom. A drawback is the computational power required, but this has been largely overcome with modern computing. While we know the functional form of individual sets of measurements follows an exponential decay (Equation 1), pooling normalized data, especially in the background normalized case, could change the underlying functional form in subtle but important ways. The flexibility of loess can potentially capture this variation, to the extent that it exists.

In locally weighted regression, low-degree curves are fitted to subsets of data and aggregated to form the final curve. Specifically, for each value of an independent variable ( $x$ ), say  $x_i$ , an estimate is made which uses some neighbors of  $x_i$ , which is defined as a parameter to the model. Neighbors closest to  $x_i$  are given more weight than those further away, according to a weighting function. A linear or quadratic curve (typically)<sup>17</sup> is then fitted to predict the dependent variable as a function of the subset of independent variable(s) using the previously determined weights combined with an ordinary least squares algorithm. The model estimation is that value. This proceeds for each value of the independent variable in the dataset. In this sense, loess is similar to moving average methods typically used for time series data.

In addition to the degree of fit (linear or quadratic), the key model parameter is the degree of smoothing, referred to in R as “span” or alpha in the call to `loess()` and by Cleveland (1979) as  $f$ . Simply, this parameter indicates how large the neighborhood used to calculate the local regression models will be. A smoothing parameter equal to one will include all the points in the dataset—as this parameter increases, the smoothness of the line increases as well.

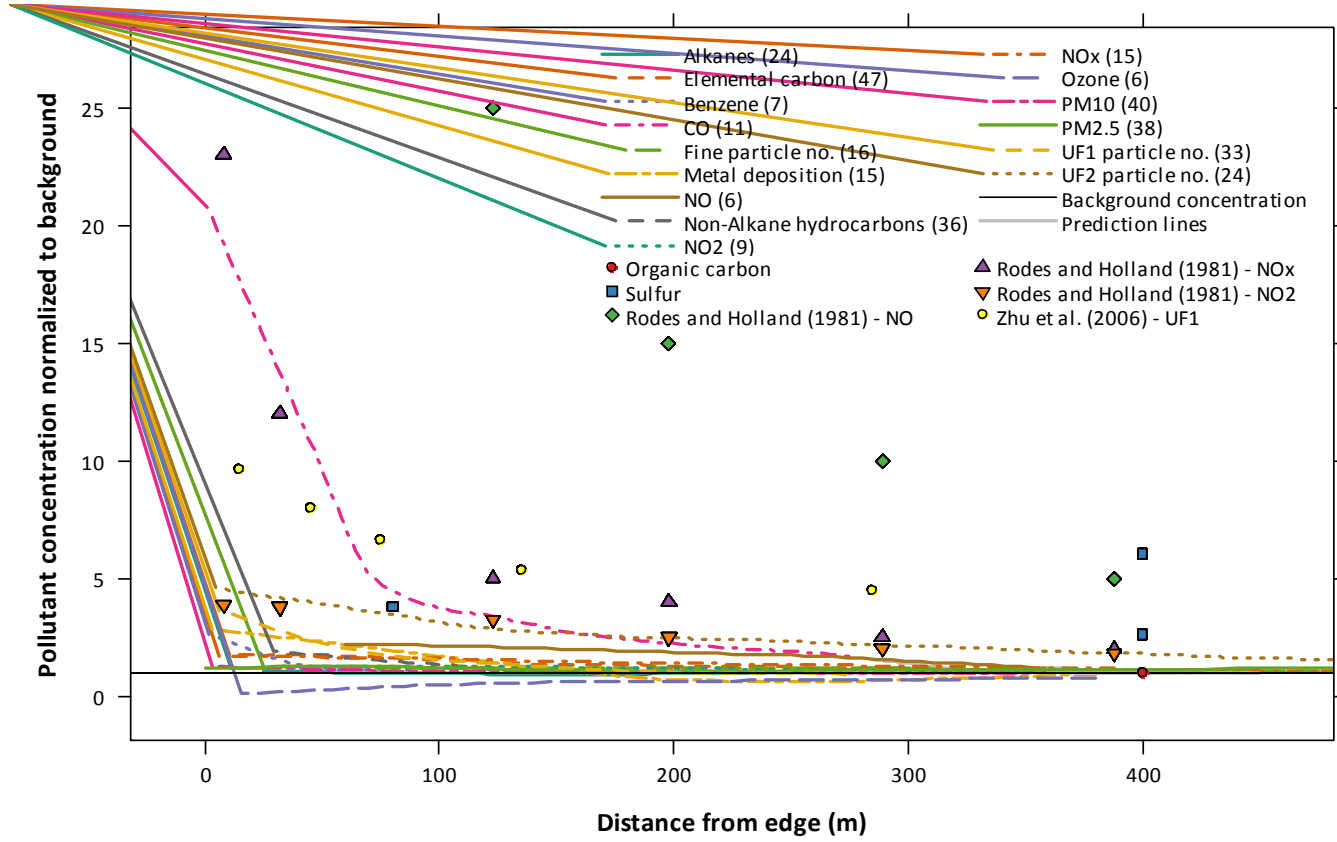
---

<sup>16</sup> Implemented in R as `loess()`, but variations on similar themes are available including `supsmu()`, `locfit()`, and others. Smoothing performed in the present study utilized only the `loess()` function.

<sup>17</sup> Linear estimations are preferred, unless the data exhibits unique features such as local minima and maxima.

Locally weighted regression has previously been applied to near roadway data by Gramotnev and Rivoski (2004) who employed the loess method. The authors did not specify the value of the smoothing parameter used, but noted that the loess regression was an efficient way to represent datasets with a small number of experimental points. In our case, the smoothing parameter was set by visual inspection—the general shape of the curve was known, namely that for pollutants which vary by distance, there should generally be a decrease from the value closest to the road and then a flattening. However, for the reasons discussed above, the smoothed curves relative to background concentrations may not behave similarly due to the preservation of variations by magnitude. Setting the smoothing parameter for all pollutants equal, an iterative procedure was employed where the parameter was increased from its default value of  $2/3$  until the majority of pollutants were amenable to smoothing. Generally, with a small sample size and large distance between observations, the smoothing function will fail. A value of 0.75 appeared to impart a reasonable shape to each curve, while the curves for organic carbon ( $n = 2$ ) and sulfur ( $n = 4$ ) were unable to be drawn even at that level of inclusion. A degree of one, indicating locally *linear* fits, was used in all cases.

The results of smoothing the background concentrations are shown in Figure 4. The top panel (a) includes the entire range of data, while the bottom panel (b) includes a closer view of a subset, where the maximum value on the ordinate is restricted to five times above background concentrations. The range on the abscissa has been constrained to 0 – 450 m. Remaining figures all have similar ranges since there is relatively little data that lies outside of that envelope. Organic carbon and sulfur are plotted as points instead of smoothed lines, due to the limited amount of data, and four other sets of measurements were removed and plotted as points as well—Rodes and Holland's (1981) data for oxides of nitrogen ( $\text{NO}$ ,  $\text{NO}_2$ , and total  $\text{NO}_x$ ) which were taken during a period with substantially different fleet characteristics than exist at the date of the majority of the studies in the database, and Zhu et al.'s (2006) data for nighttime concentrations of UF1 particle number which were taken under such different meteorological conditions than every other study that including their results leads to counterintuitive locally weighted regressions. No such exclusion was necessary in the edge normalized case.



(a)



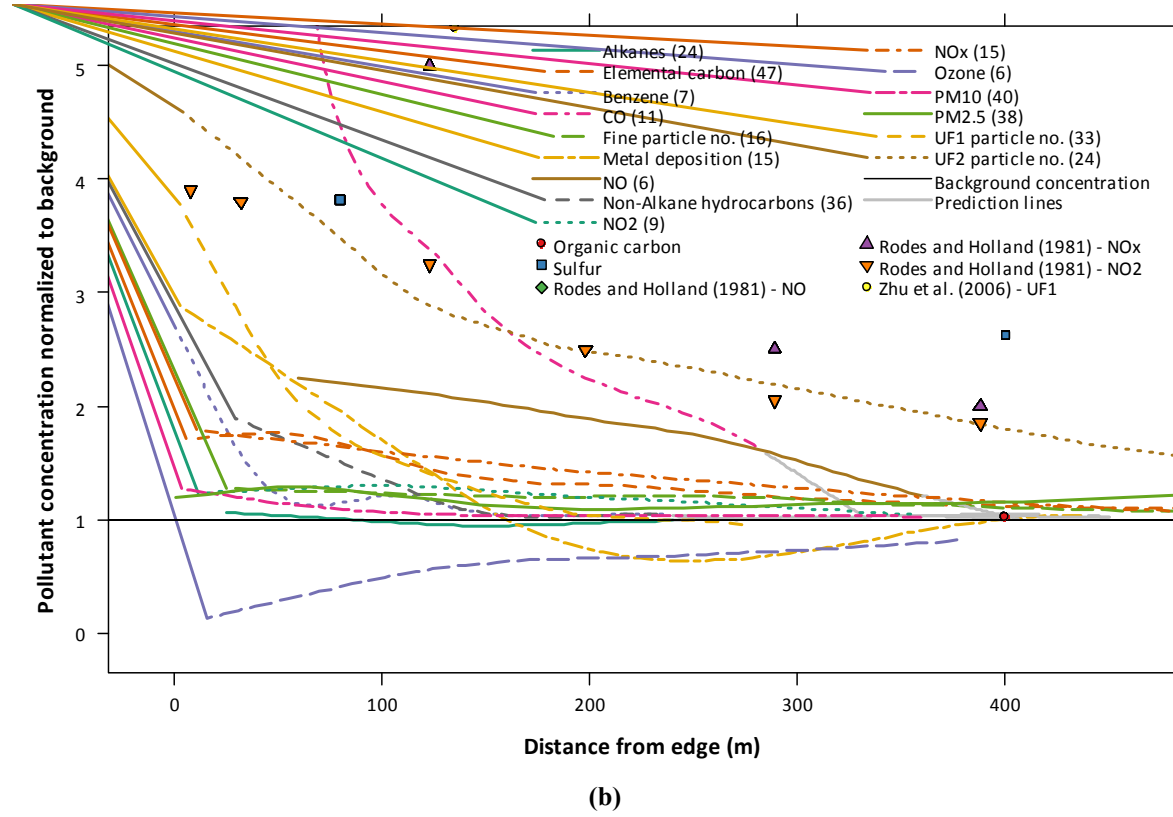


Figure 4: Local regression of background normalized concentrations on distance (a) for the entire range of data; (b) for a subset of the data limited to five times above background concentrations, to show increased detail.

Note: The single difference between (a) and (b) is the scale on the ordinate. The horizontal thin solid black line indicates background concentration: a value of unity on the ordinate. A loess smoother ( $\alpha = 0.75$ , degree = 1) is fitted to each pollutant. Plotted points are shown for those data not suitable for smoothing, or those data collected under unique conditions (e.g., nighttime) which rendered them less comparable to other data (see text for further discussion of excluded data). Organic carbon and sulfur were too sparse to fit a smoother to. The sample size,  $n$ , used to construct the smoother is given in parentheses after the pollutant type in the legend. The Zhu et al. (2006) data shown is for nighttime concentrations and is plotted separately using individual points since different meteorological conditions prevail at night than during the day. That study is the only one from the literature we reviewed that took nighttime measurements. The Rodes and Holland (1981) data for oxides of nitrogen were also plotted separately using points. Its inclusion was found to substantially bias the curves upwards, since it was from a time period with substantially different fleet characteristics than exist for the rest of the studies.

Since the local regression shares similarities with moving average methods, the figure should be thought of as representing the best estimate arrived at as the result of pooling the insights of the near road studies undertaken to date. In interpreting the figures that follow, the reader should note that there is generally far more data in the first 150 m than beyond, and lines constructed with larger sample sizes (shown in parentheses in Figure 4's legend) are likely more robust than those created from smaller sample sizes.

In Figure 4a, the high values of CO in general, and of nitrogen oxides from the Rodes and Holland (1981) study, extend the maximum values on the ordinate so high that further distinctions by pollutant are made difficult. The data show that CO concentrations are elevated approximately 20 times above background concentration in the vicinity of the road. Looking at Figure 4b, which restricts the maximum value on the ordinate to five times above background concentrations, more distinctions can be made. Note however that one of the limitations of the background normalization approach is that it mixes study results that have different background concentrations. The variability in starting background concentration can affect the shape of the curve, particularly close to the road. For example, several of the smoothed lines (most notably elemental carbon and its surrogates) are concave or have an inflection point when they should be convex. This is likely caused by the inclusion of points which exert opposing forces on the smoother. If an exponential curve were instead fitted to these data, such behavior would be obscured, or a concave exponential curve would result. The smoother lets small features of the data affect how the curve is drawn. Other pollutants show intuitive results under local regression: ultrafine particle numbers, benzene, and non-alkane hydrocarbons all begin at high concentrations, decrease approximately exponentially, and reach a background concentration at some distance downwind. The variation in observations was low enough to result in a line which conformed to expectation for these pollutants. This is a shortcoming of background normalization—it requires more homogeneity among the data to produce the expected line than will be shown to be the case for edge normalization.

Several pollutants show noticeably flat smoothed lines, with values of concentration in the vicinity of the background value. These include  $PM_{2.5}$ ,  $PM_{10}$ , fine particle number (including particles with aerodynamic diameter greater than 300 nm), alkanes, and  $NO_2$ . The first four have all been reported in the literature as showing no or little trend near the road.  $NO_2$  is formed secondarily in the atmosphere from the oxidation of NO (which is directly emitted by

automobiles), and should show a relationship with distance (similar to the NO<sub>2</sub> data plotted from the excluded study in Figure 4b), but it should be more gradual than the steep declines seen for ultrafines, for example. The reasons for this behavior have been explored elsewhere (see, e.g., Zhou and Levy 2007), but they deal mostly with the relative contributions of chemical reactivity and dispersion in the atmospheric transport and fate of a given pollutant. In general, NO<sub>2</sub> is formed near roads and ultrafines are removed as particles grow due to various aerosol dynamics (Zhang et al. 2004). However, this could also be a limitation of the available studies for NO<sub>2</sub> with measured background values (only two studies, with n = 9 total, one of which definitely employed averaging times which exceeded the peak traffic period (Kim et al. 2004) which would tend to diminish differences between background and near road values). NO<sub>2</sub> does show variation with distance when using edge normalized values, as will be shown below.

A middle group consisting of elemental carbon and its surrogates and NO<sub>x</sub> appears to be elevated near the roadway and then to decrease gradually to background values. Both pollutants are elevated between 1.7 and 1.8 times background near the roadway and decay by 400 m to background levels. NO lacks data in the immediate vicinity of roads, but appears to be trending upwards the closer to the road the observations occur; it is elevated above total NO<sub>x</sub> and NO<sub>2</sub> as should be expected, since NO is directly emitted by automobiles. An additional limitation of the NO observations is its small sample size (n = 6). Based on this limited data, it appears to reach background levels by 400 m. A third category is elevated above background levels, but decreases rapidly—non-alkane hydrocarbons are elevated 2.2 times above background and decrease to a normalized value of one by 150 m; benzene is elevated approximately three times above background values and decreases to background by 75 m; metal deposition is elevated three times above background near road, but reaches background by 175 m and shows a curious dip between 175 and 400 m which is due to one errant measurement which pulls the smoother below background concentrations; UF1 particle number (counting particles greater than 3 nm in aerodynamic diameter) is elevated four times above background near road, drops to twice background by 75 m, and fully reaches background values by 200 m; UF2 particle number (counting particles greater than 15 nm in aerodynamic diameter) is elevated almost five times above background near road, decreases to three times above background by 115 m, and is still almost twice background at 400 m; CO decreases from its high of 20 times background to five times by 90 m and likely reaches background at approximately 330 m; Ozone shows an

increasing trend which is far below background values near the road, and gradually approaches background by 400 m from the edge. Ozone is the only pollutant to exhibit this behavior; the ozone values are consistent with expected near road titration (removal) due to interaction with direct vehicle emissions of NO. The discussion that follows generally ignores the ozone data and refers only to those pollutants which exhibit decreasing trends with distance. These results suggest a range of approximately 75 – 400 m to reach background concentrations, depending on the pollutant.

Particle number concentrations can be expected to increase as the lower end of the size range measured is decreased. Thus, the result that UF2 particle number exceeds the UF1 particle number is curious, but can be explained by recalling the limitations of background normalization—that it tends to retain between study differences in magnitude. One study which measured UF2 particle number was Hitchins et al. (2000). The authors did not take background measurements, but did report concentrations when the wind direction was from the receptors to the road. Many other studies report measurements under such conditions as representative of background concentrations since under these wind conditions, the detected values are devoid of roadway influence (see, e.g., Gidhagen et al. 2004). Additionally, the study reported concentration values under several different wind speed scenarios. The highest values of UF2 particle number were reported at the lowest wind speed, but the background value was given for conditions with a higher wind speed. This highlights the difficulties associated with pooling data from studies that used different measurement and reporting protocols.

All pollutants except for UF2 particle number reach background by approximately 400 m. Since the smoother represents a *local* regression, it is not necessarily advisable to use the lines for extrapolation, however, the gray lines in Figure 4 were constructed by extending the final smoothed line segment beyond the range of observations and are intended to be used as a visual guide only, to highlight the convergence at 400 m and not as representative of actual values. The high variability shown in the first 100 m drops off by 400 m, even considering the between-study differences in methods and traffic characteristics, which is notable considering the wide variation in data and the inherent problems with this normalization method. The curves all agree (ignoring ozone) with the intuitive notion that concentrations of certain pollutants are elevated near roadways, and decrease as the distance increases. Preliminary testing of the unnormalized

data using loess models showed radical departures from these patterns (not shown), which supports the use of normalization procedures.

We now return to the original data points and group them by variation with distance. Figure 5 shows the data used to construct the smoothed lines separated into its own panels by pollutant and grouped according to variation by distance. The scale of the axes is determined by the range of data in each panel, much like Figure 2. Unconstrained axes are necessary with background normalization to show the full extent of the data, since they are not constrained to lie within the same range, as in the edge normalized case.

Wide variation is evident among pollutants which vary by distance, but those that do not are grouped relatively close together, which shows that their range of elevations above background is far lower, supporting the categorization of them as non-varying. One exception can be found among the  $PM_{10}$  measurements in the immediate vicinity of the road which appear to be elevated between two and three times above background. Several of these high observations are from a study conducted in Macao, China (Wu et al. 2003) in which a low background measurement tended to inflate the near road values. On net, these elevated values of  $PM_{10}$  were not enough to affect the smoother since the majority of the data lies closer to a normalized value of one, but if a trend exists, its magnitude is presented in the statistical analysis below.

As previously mentioned, variations in magnitude within each panel are still visible. The NO,  $NO_2$ , and  $NO_x$  panels highlight the effects of fleet turnover and year of study on observed concentrations—the green diamonds in those panels are from the Rodes and Holland (1981) study which were excluded from smoothing. In the case of NO and  $NO_2$  the excluded points clearly lie outside of the range indicated by the other points in the panel. However, the  $NO_x$  panel shows two blue diamonds in the same concentration range as the excluded points, highlighting another example of a low background concentration. The blue diamonds originate from a study (Gidhagen et al. 2004) that had lower background concentrations than the others in the database by an order of magnitude since it was conducted in a rural area, whereas the majority of the other studies were urban. This disparity inflates the near road increment relative to the others, when normalizing to the background concentration.

The UF1 and UF2 particle number panels show wide variation and reflect the differences in observed background concentrations, traffic, and meteorological conditions observed.

However, those measurements derived from a single study (e.g., metal deposition and non-alkane hydrocarbons) show relatively good agreement with this normalization method, supporting its use in situations where variations in study parameters are consistent across measurement sets. Benzene and CO also appear to show a consistent pattern, even though the data originate from several different studies. In general however, it appears that, using background normalization, the degree to which consistent results can be obtained from pooled data will vary substantially especially when data has been collected across long time periods and in rural vs. urban locations. One identified problem is the difference in what constitutes background measurements from site-to-site.

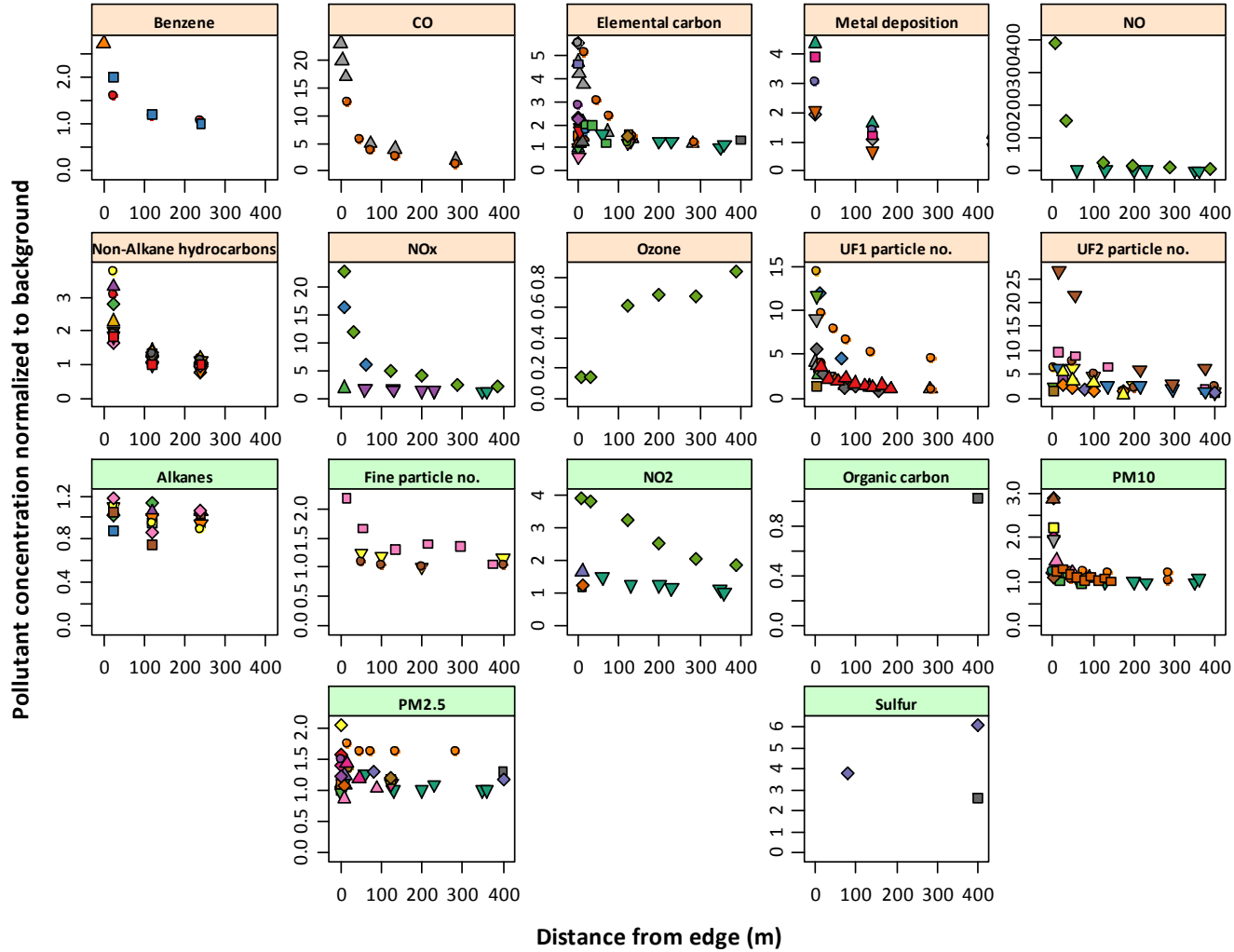


Figure 5: Background normalized near road air pollution data plotted in separate panels.

Note: Different plotting characters within each panel represent unique sets of measurements either from a single study or from different studies. The top two rows with the orange strip show pollutants whose concentrations vary by distance. The bottom two rows with green strips do not show clear distance from road trends and are comparatively flat when compared to those that do vary. The plotting ranges are determined by the range of data; all data for each pollutant is plotted in each panel.

#### 5.2.2.1.2 *Analysis of variance: Overview Discussion*

Statistical significance is rarely discussed in the near road literature. Roorda-Knape et al. (1998) and Smargiassi et al. (2005) used paired t-tests to compare observed concentrations to a reference group typically located closest to the roadway. However, Zhu et al. (2004) note that particle number concentrations (in addition to other meteorological and traffic measurements) in summer and winter did not meet the normality criteria for a t-test. The authors instead conducted a nonparametric Mann-Whitney test which replaces the observation with its ranking without regard to grouping, and compares the sum of the ranks in each group. This test does not require an assumption of normality. In addition to paired t-tests, Sabin et al. (2006) used an ANOVA to test differences in the dry deposition rates of metals between downwind locations. ANOVA additionally requires the assumptions of equal variance among groups and independence among sampled data. The underlying assumptions of the ANOVA and the t-test were not discussed by the three studies which conducted them. Violation of basic assumptions is another possible reason for the lack of significance assessments in the literature. Below, we discuss some of the issues with assessing the statistical assumptions of ANOVA with near road concentrations, using our dataset as an example.

ANOVA compares means of groups where each group might vary in some way from the others, more than would be expected by chance.<sup>18</sup> The null hypothesis is that there is no difference in group means. To determine the veracity of this hypothesis, the method calculates the ratio of two independent estimates of population variance. The numerator includes an estimate based on the chance variability and the effects of the grouping. The denominator is an estimate of the variance based only on chance. If the ratio is large, it is possible that the grouping has a significant effect on the means, and that we can reject the null hypothesis. The closer the ratio to one, the more likely we are to fail to reject the null hypothesis. If the variance across the groups is homogeneous, the samples are random and independent, and the population is normally distributed, the ratio of the two estimates of variance is  $F$  distributed. The  $F$  statistic that results from the ratio calculation is referred to as the “omnibus”  $F$  and the probability of observing it can be calculated. If the probability,  $\alpha$ , is small enough, we can safely reject the null hypothesis

---

<sup>18</sup> This discussion of ANOVA and the limitations of our dataset draws heavily from Myers and Well (1991), specifically their chapters three and four.



with confidence equal to  $1 - \alpha$  (i.e. if  $\alpha < 0.05$  then we reject the null hypothesis with 95 percent confidence).

For the present dataset, ANOVA can illuminate the magnitude and significance of the decrease in concentration by distance group. A categorical grouping variable is required to perform the analysis. Inspection of Figure 4 led to the selection of three groups for comparison: 0 – 80 m, 80 – 120 m, and > 120 m. Clearly, different atmospheric transport processes affect the decay to background concentrations for different pollutants, but two types appear to be evident: those that are relatively flat and decline gradually, and others that decline rapidly before flattening. Dividing the initial decline into two separate bins allows the model to elicit both types of behavior. Some pollutants have no data in the second bin, but this simply reduces the test to a comparison between the first and third groups. The null hypothesis in this case is that there is no difference in mean observed concentrations between distance bins.

Violation of any of the ANOVA assumptions, however, can lead to errors in assessing the null hypothesis. Type I errors result when the null hypothesis is true, but it is rejected (false positive). Type II errors arise when we fail to reject a false null hypothesis (false negative). The researcher sets the desired Type I error threshold, but it is only accurate if the ANOVA assumptions are met. The rate of Type II error can be controlled by selecting a statistical test appropriate for the population distribution. There are no procedures that can unequivocally assess adherence to the underlying assumptions of ANOVA, but we discuss each violation and possible consequences below.

Violations of the assumption of normality are generally not problematic unless the sample size in each group,  $n$ , is small and the deviations from normality are severe. The values for sample size presented for the background and edge normalized ANOVAs are generally small (Table 4, and Table 5 below). To determine the distribution of observations in the population of pollutant concentrations near a road, assume that a theoretical series of measurements was taken continuously at randomly sampled distances from 0 – 500 m from the roadway edge. The pollutant under observation varies with distance according to a negative exponential decay (Equation 1). The distribution of observations over this distance range would be positively skewed since there would be a higher density of lower observed concentrations, but as the distance considered decreased in size from the entire range to a single point, the remaining variation would arise from the normally distributed error term, rendering normal the distribution

of observations at that point. However, we are sampling over the entire range, and thus have a skewed population. A statistical test that does not presuppose a normally distributed population might provide more statistical power (confidence in rejecting a false null hypothesis) (Myers and Well 1991, p. 103-4) in this case. This possibility is discussed after the assumption violations are considered further.

Inhomogeneous variances result in an increased rate of Type I error (more likely to reject a true null hypothesis and thus to detect an effect of grouping where none exists), but the increases are generally small unless the sample sizes within each cell are imbalanced. Inspection of Table 4 shows that most pollutants have imbalanced cells meaning that the number of observations at each distance range is not equal for a given pollutant..<sup>19</sup> The effect on the Type I error rate depends on whether there is an association with variance and group size. If larger sample sizes have larger variances, the *F* statistic will be too small, and if small sample sizes have the larger variances, the *F* statistic will be inflated. We expect homogeneity of variances to increase for edge relative to background normalization since the data are constrained to lie within a smaller range and the variations due to magnitude are better controlled for, as previously discussed.

The final assumption to consider is independence of observations. As noted by Gilbert (1987), pollution monitoring data is frequently correlated both in time and in space, in the sense that high (or low) values are likely to be taken close to other high (or low) values. If the data were all for a single distance from road point, this would be a non-issue; however, our dataset likely includes dependence within sets of measurements. But, we have pooled data from many different studies together—this means that the higher the *n*, the more that dependence has been reduced within that pollutant. To the extent that this assumption is violated, the Type I error rate will be inflated. Related to this assumption is an additional one that the data included in each group is randomly sampled, so that its sample parameters are unbiased estimators of the population's. An examination of Figure 5 seems to indicate that no distance range was systematically excluded from analysis, but there is more data in the first 150 m than further away (Figure 1). The effect of having more data closer to the road is unknown. If the variance is high

---

<sup>19</sup> The imbalance could be remedied by letting the data determine the size of the ranges studied, but given that there is a preponderance of data near the roadway, ranges determined in this manner would not cover the range over which a trend exists. Further, although the cells would be balanced, they would differ by pollutant which would render comparisons difficult.

in those groups then the Type I error rate would be lowered, but it is likely that the normalization techniques have reduced the variance within each group considerably.<sup>20</sup> Two possible strategies to reduce this error are further pooling of the data, or the generation of additional data by simulation methods. We do not consider these strategies here.

We noted above that a test which does not presuppose a normal distribution might provide more statistical power (reducing the likelihood of accepting a false null hypothesis) in the case of a skewed or heavy-tailed distribution. An approximately equivalent method to the ANOVA that fits this description is called the Kruskal-Wallis test. The Kruskal-Wallis test replaces each measurement with its rank irrespective of group membership. The ranks are then placed into groups and their means are compared. The null hypothesis is that the populations from which each group was drawn have equivalent means (Gilbert 1987). The populations need not be normally distributed, but each should have a similarly shaped distribution, and the cell sizes need not be equal. Inhomogeneous population variances affect the Kruskal-Wallis test as they affect the ANOVA, but the former test should be somewhat less sensitive to these.<sup>21</sup>

For each pollutant, we support the ANOVA results with the nonparametric Kruskal-Wallis test and discuss differences when they arise. In light of the limitations of this data, the ANOVA results presented should be considered preliminary pending further data acquisition and analysis.

Even pollutants which appear to vary little with distance could have a statistically significant trend, but emphasis then shifts to the magnitude. If it is significant but small, a roadway influence can be inferred, but it is likely that the roadway is not the most important source of the pollutant and mitigation efforts and the design of epidemiological studies might be better focused elsewhere.

---

<sup>20</sup> Normalization is also required for the validation of the ANOVA model. A one-way ANOVA posits that one factor is most important to the outcome. In this case we propose that distance group is the only factor that predicts concentration. From the literature review, we know that there are many factors that affect the concentrations, but normalization is expected to reduce that variation such that distance is the only independent variable remaining. This should be the case to a much greater extent in the case of edge normalization since we have previously demonstrated that wide variation can still remain even after normalizing to background concentrations.

<sup>21</sup> To test the effect of small sample sizes and inhomogeneous variances on the results of the Kruskal-Wallis test, the data were split into two instead of three groups (0 – 120 m, > 120 m) and both the Mann-Whitney and Kruskal-Wallis tests were performed on the data. The case of two groups had less heterogeneous variance from a visual inspection of box-and-whisker plots of the data (not shown). The conclusions for significance under each normalization did not change substantively. In the background normalization case, NO<sub>2</sub> showed a significant difference under background normalization and benzene showed significance under edge normalization.

The results of the ANOVA for background normalized concentrations are shown in Table 4. Those pollutants excluded from smoothing in Figure 4 and plotted instead as points are also excluded from this analysis. A linear model was fit to the data in R using `lm()`, and the ANOVA table was extracted using `anova()`. The omnibus  $F$  is reported along with the degrees of freedom from the ANOVA table. The significant regression coefficients are reported from each model, if the omnibus  $F$  or Kruskal-Wallis  $p$ -value indicated significance at the 95 percent level. The sample size,  $n$ , is also shown for each cell regardless of significance. Missing data in the table indicates that no concentrations in that size range for that pollutant were recorded or that results were insignificant. The coefficient in the distance range 0 – 80 m was reported as the intercept for the model, and so its significance is not included in Table 4 since it would only indicate the significance of difference from zero.

#### **5.2.2.1.3 Analysis of variance: Results**

All regression coefficients reported in Table 4 can be interpreted as the mean values for the pollutant in the distance range and coefficients in the second and third distance bin are mean changes relative to the first bin. All of the coefficients in the second and third bin are negative (except for ozone), as expected for this air pollution data. In general, the ANOVA shows a significant difference between groups for the pollutants identified as varying with distance. Mean values of elevations above background in the first 80 m from the road range from 2.1 for benzene to 12.4 for CO. Surprisingly, NO does not seem to vary significantly with distance from road. However, its sample size is very small, and is all drawn from one dataset with a long averaging time, as previously mentioned. Both of these issues could be obscuring the detection of differences by distance for NO. The Kruskal-Wallis test also supports the hypothesis that there is no difference between the groupings of NO as well as NO<sub>2</sub>, with large  $p$ -values in both cases.

PM<sub>10</sub> is the only other pollutant for which the omnibus  $F$  from the ANOVA differs with the  $p$ -value from Kruskal-Wallis. This implies a possibly significant relationship with distance from road for PM<sub>10</sub>; its magnitude would be the lowest among all pollutants. PM<sub>10</sub> is on average 1.4 times above background concentrations, and declines by an average of approximately 0.4 from the near road value beyond 120 m, representing a 30 percent decrease. Benzene shows the next smallest significant increment above background at 2.1, declining to 1.1 past 120 m, a drop of 48 percent. Each of the other pollutants which vary by distance both begin at a higher above-background increment than PM<sub>10</sub> and decrease more sharply. Thus, the relationship of PM<sub>10</sub> with

distance appears to be small if existent. PM<sub>2.5</sub>, fine particle number, and sulfur do not vary by distance, and alkanes show a very small statistically significant relationship with distance bin, decreasing by eight percent over the distance range greater than 120 m from the road. The generally good agreement of the Kruskal-Wallis and ANOVA results lends support to the use of either method for the detection of magnitude difference by distance groups in near road air pollution, but violation of underlying ANOVA assumptions will need to be investigated further to assess differences between the parametric and nonparametric tests.

**Table 4: ANOVA Results for Background Normalized Pollutant Concentrations**

Pollutant	ANOVA		Kruskal-Wallis $p$ -value <sup>a</sup>	0 – 80 m		80 – 120 m		> 120 m	
	Df	$F$		Coeff. <sup>b</sup>	n	Coeff. <sup>b</sup>	n	Coeff. <sup>b</sup>	n
Benzene	(1,5)	12.44*	0.034	2.10	3			-0.992*	4
CO	(1,9)	6.14*	0.014	12.41	7			-9.97*	4
Elemental carbon and surrogates	(1,45)	9.23**	< 0.001	2.40	32			-1.111**	15
Metal deposition	(1,13)	30.8***	0.0022	3.06	5			-1.947***	10
Non-Alkane hydrocarbons	(1,34)	69.6***	< 0.001	2.37	12			-1.250***	24
NO <sub>x</sub>	(1,13)	4.80*	0.0022	5.55	5			-4.26*	10
Ozone	(1,4)	58.4**	0.0603	0.1429	2			0.561**	4
UF1 particle no.	(2,30)	3.86*	< 0.001	4.50	22		2	-3.32*	9
UF2 particle no.	(2,44)	5.71**	< 0.001	6.88	18	-3.99*	7	-4.58**	22
Alkanes	(1,22)	4.31*	0.043	1.053	8			-0.0817*	16
Fine particle no.	(2,13)	3.49	0.140		4		2		10
NO	(1,4)	1.852	0.132		1				5
NO <sub>2</sub>	(1,7)	4.63	0.288		6				9
PM <sub>10</sub>	(2,37)	3.13	0.00237	1.432	26		3	-0.388*	11
PM <sub>2.5</sub>	(2,35)	0.472	0.686		22		2		14
Sulfur	(1,2)	0.0314	0.655				1		3

**Note:** The heavy black line separating UF2 particle number from alkanes indicates those pollutants determined to vary with distance (above the line) and those that do not or do so with low magnitude (below the line). Statistical significance is indicated as follows: \*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$ . Missing values in the table indicate insignificant results as judged by the omnibus  $F$  and the Kruskal-Wallis  $p$ -value, or no data, evidenced by a blank n for the cell.

<sup>a</sup>The Kruskal-Wallis  $p$ -value was determined using the `kruskal.test()` function in R which performs the non-parametric Kruskal-Wallis test whose null hypothesis is that there is no difference in the mean ranks of the groups.

<sup>b</sup>Refers to the regression coefficients extracted from a linear regression of normalized concentrations on distance bin. The coefficient in the 0 – 80 m bin was the model intercept which represents the mean normalized value in that range, while the coefficients in the other two distance bins represent mean changes relative to the first bin.

### 5.2.3 Results: Near roadway concentrations normalized to roadway edge

This section completes the same analysis as that for the concentrations normalized to background, but brings additional and somewhat different data to bear on the problems. Background values are not necessary for edge normalization; however, to the extent that the values show evidence of leveling off, that percentage value will approximate a background (or near-background) concentration—the point at which the leveling begins indicates the range of roadway influence for that pollutant.<sup>22</sup>

Figure 6 shows the 480 distance/concentration pairs for data normalized to the edge of the roadway grouped both by pollutant and “class,” as previously defined. With the exception of ozone and sulfur, Figure 6 indicates a generally decreasing trend as distance from road increases, although a wide band of concentrations is still visible even out to 450 m. The reasons for this are clarified when examining the class groupings in the right panel. Most of the measurements that do not trend downward as distance increases are the PM mass concentrations composed of the criteria pollutants PM<sub>10</sub> and PM<sub>2.5</sub>, identified above as having very small near road elevations. Particle numbers (identified as “Number” in the right panel of Figure 6), on the other hand, generally trend downwards in the right panel, while plotting all gases together (identified as “Gas” in the right panel of Figure 6) shows no consistent trend. Grouping “Other” pollutants together also shows no consistent pattern.<sup>23</sup> Additionally, far more data for NO is available in this edge normalized dataset than was available for background normalization, and its relationship to distance is discernable in the left panel of Figure 6, showing very clear elevations near the roadway and sharp declines.

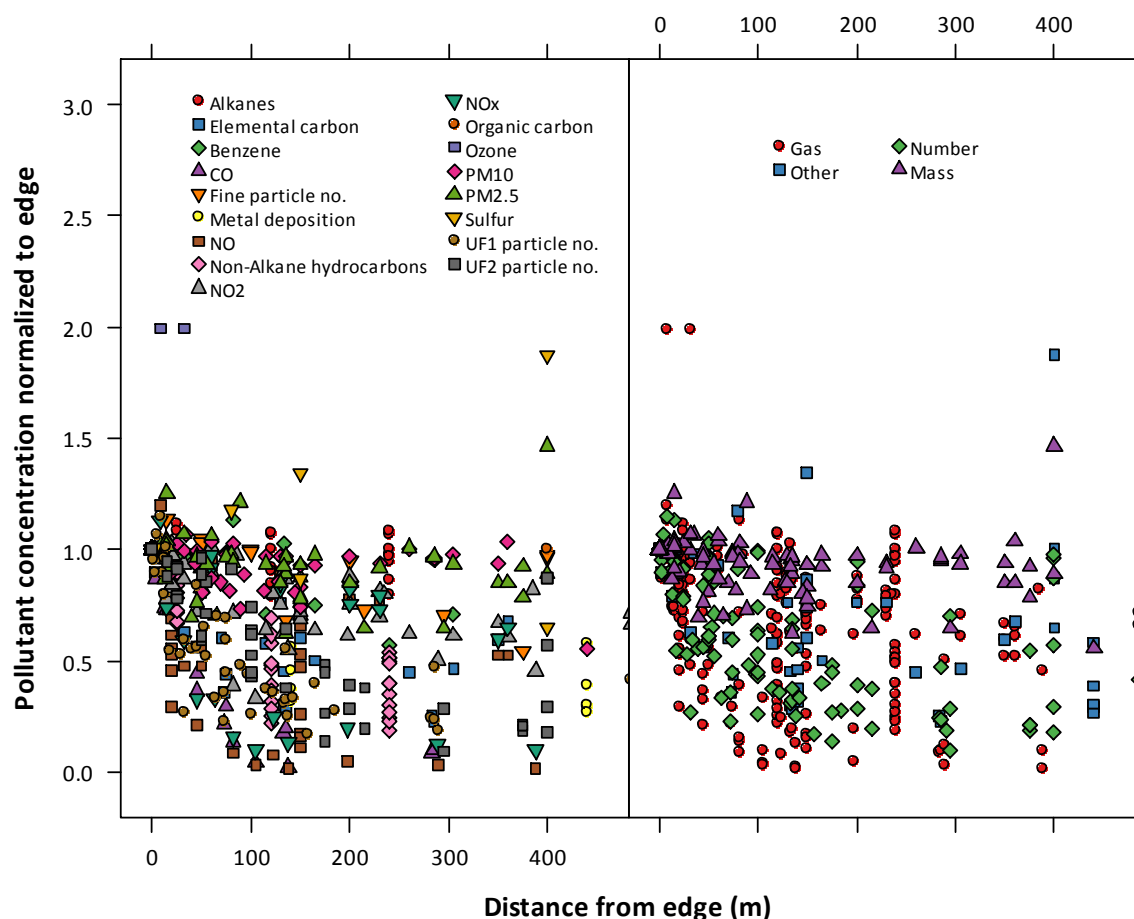
Another property of edge normalization is that the data are constrained to lie in a much narrower range than with background normalization. The vast majority of points lie below one on the ordinate, and the only pollutants that oscillate above and below that value are those which show increasing, flatter, or ambiguous trends with distance—ozone, sulfur, alkanes, PM<sub>2.5</sub>, and PM<sub>10</sub>. Note that this is not dependent on how the edge concentration was predicted. If a linear fit was used to predict the edge concentration for a pollutant that actually exhibited exponential decay behavior it would tend to shift the curve past one in the vicinity of the road, but it would

---

<sup>22</sup> Recall the discussion in section 5.2.1 that hypothesized a “perfect” near road dataset.

<sup>23</sup> This does not mean that it is not possible for any one gas to trend downwards, just that this pattern is not evident when viewing all gaseous concentrations together.

still exhibit an exponential decay shape, since normalization, in this case, only affects the magnitude of concentration, not the shape of the curve. Exponential fits were used to determine the edge concentrations for elemental carbon, benzene, CO, metal deposition, non-alkane hydrocarbons, NO<sub>x</sub>, ozone, and UF1/UF2 particle number since background normalization showed that they varied exponentially with distance. Due to the lack of data for sulfur for background normalization, an exponential decay model was used for its edge values. NO and NO<sub>2</sub> edge values were also estimated using an exponential fit, based on findings from the literature cited in Appendix B. All of the other pollutants that showed little variation with distance (PM<sub>10</sub>, PM<sub>2.5</sub>, fine particle number, and alkanes) used linear regression to estimate an edge concentration.



**Figure 6: Edge normalized near road air pollution data.**

**Note:** The left panel shows all pollutants by pollutant type, while the right panel collapses pollutants into their “class” as previously defined.

### 5.2.3.1 Trend analysis

As in the background normalized case, below we perform a local regression and an ANOVA on the data normalized to edge concentrations, utilizing identical methods to those previously described.

#### 5.2.3.1.1 *Local regression*

The smoothed lines for each pollutant normalized to the roadway edge are shown in Figure 7. Values on the ordinate represent a percentage decrease from the edge of road value. There appears to be a wide variation amongst pollutants in this plot, but far less than the background normalized case since each observation is constrained to lie within zero and one (with the exception of ozone). Some pollutants show no clear trends with distance, some show very clear trends, and others lie in between. Clearly,  $PM_{10}$ ,  $PM_{2.5}$ , alkanes, and the fine particle number show no or ambiguous trends with distance, as was the case with the background normalized data.

For the group of pollutants that vary by distance, the percentage of the near road high at which leveling occurs represents the smoother's estimation of background concentration of that pollutant since the roadway influence has dropped to approximately zero when no further changes occur in the smoothed curve. Some variation is still possible, especially in the more sparsely populated regions of the smoother, since a single point can influence the plotted curve in the absence of surrounding data.  $NO_2$  and benzene show some decreasing trend, decreasing by 30 percent and 40 percent from their edge of road high by 300 m, respectively. The trend for  $NO_2$  agrees more closely with knowledge from the literature on its atmospheric transport and fate than was shown for background normalization. The final group of pollutants all exhibit very sharp declines within the first 100 – 150 m before leveling off. Concentrations of elemental carbon and its surrogates decrease by approximately 40 percent within the first 90 m from the roadway edge before flattening;  $NO_x$  trends similarly, decreasing by 50 percent within 100 m; non-alkane hydrocarbons and UF1 particle number both decrease by 65 percent within 150 m; finally, NO and CO show the greatest declines. NO drops by 60 percent from its near road high within 80 m, and CO drops by 90 percent within 120 m. A range of between 80 to 150 m appears to describe the evolution to background concentrations for most pollutants.

Larger percentage declines indicate a greater contribution of the roadway above background levels—the reciprocal of the percent of the near road high (i.e. 1 – percent decrease)



at which the smoother levels off represents the multiplier at the edge of road above background concentrations (see Table 6). Some leveling is apparent for benzene (60 – 70 percent of near road high) and  $\text{NO}_2$  (67 percent of near road high) but their curves still show a downward trend, indicating that background concentrations are not yet reached over the plotted distance range. Leveling is most apparent for elemental carbon and surrogates (60 percent of near road high),  $\text{NO}_x$  (54 percent of near road high), non-Alkane hydrocarbons (35 percent of near road high), NO (33 percent of near road high), UF1 particle number (25 – 30 percent of near road high), and CO (< 10 percent of near road high). Some variation can be seen between 300 and 400 m specifically for  $\text{NO}_x$  and UF1 particle number. At this point the data becomes sparse (Figure 1a) and the smoother has fewer points in close vicinity to construct its fitted values. UF2 particle number shows substantial declines over the entire distance range, but does not appear to level. Possible reasons for this will be explored below.

UF2 particle number does not appear to follow the trends of any of the other pollutants. It steadily decreases even to 400 m, reaching 55 percent of its near road value by 150 m. Ozone shows an increasing trend, which rapidly grows to outpace all of the other pollutants due to its inversion of the typical negative exponential decay. Sulfur and organic carbon are not amenable to smoothing due to a paucity of data, while metal deposition did not function with the loess smoother using the parameters specified for the rest of the dataset, likely due to a lack of data. Points are instead plotted for these pollutants.

As an additional check for validity, the trends for oxides of nitrogen should support the physical processes involved in their formation. Since NO is directly emitted by automobiles, we expect that it would decrease most quickly as distance from road increases, followed by  $\text{NO}_x$  which is the sum of NO and  $\text{NO}_2$ . Finally,  $\text{NO}_2$  should decrease more slowly than NO since it is not directly emitted by automobiles but formed via the oxidation of NO, a secondary process. Indeed, this pattern holds, with  $\text{NO}_2$  showing a much shallower decline than total  $\text{NO}_x$  and NO.

Since Figure 7 does not show individual points for the smoothed lines, it is unclear if edge normalization has reduced the wide variation in observations when background concentrations were used to normalize. Measurements should group more closely simply because they are constrained largely to lie between zero and one (except for ozone concentrations which increase rapidly above the value of one assigned to the roadway edge). To investigate this question, Figure 8 shows each pollutant normalized to the edge concentrations, plotted within its

own panel. Measurements above one are possible for other pollutants as well, since the nonlinear model used to estimate the edge concentration could under-predict that value.

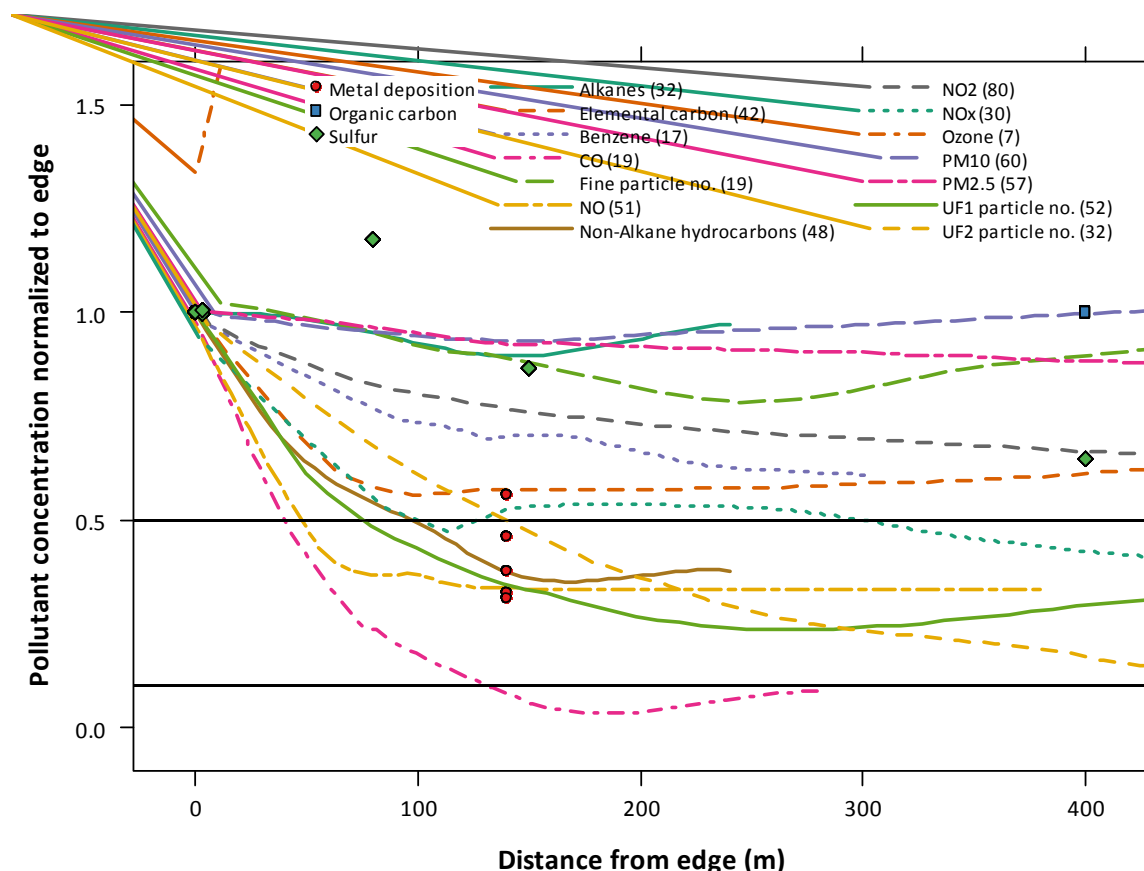


Figure 7: Local regression of edge normalized concentrations on distance.

Note: The horizontal thin solid black lines indicate two thresholds—0.1, and 0.5—which show a reduction from the edge of road concentration of 90 percent and 50 percent, respectively. A loess smoother ( $\alpha = 0.70$ , degree = 1) is fitted to each pollutant, and plotted points are shown for those data not suitable for smoothing. Organic carbon and sulfur were too sparse to smooth. Smoothing metal deposition at this value of the smoothing parameter failed in R. The sample size,  $n$ , from which the smoother is calculated is given in parentheses after the pollutant type in the legend. This value includes the additional point at the edge which was estimated for each individual dataset.

In general, the concentration data in Figure 8 is grouped more closely than the background normalized data. Elemental carbon and surrogates show agreement, despite originating from several different measurement techniques, benzene shows some disagreement, but this is likely due to the data from Roorda-Knape et al. (1998) where a high background concentration obscured the benzene trend.<sup>24</sup> Oxides of nitrogen show some interesting variation:

<sup>24</sup> In order for this to occur, the background concentration would have to be so high that it obscured the shape of the decay curve. In general, the edge normalization methods used here should detect any trend, however small. This was the case in the study by Roorda-Knape et al. (1998) which showed an inverted “U” shape for the benzene trend (grey squares in the benzene panel in Figure 8).

NO and NO<sub>x</sub> show a relatively wide disparity and two apparent patterns of decrease (one linear and one exponential, both originating from different sets of measurements as indicated by the different plotting characters associated with each) while NO<sub>2</sub> appears to show a clear decrease. One possible explanation for this disparity is the averaging time employed by each study. Flatter trends should result from longer averaging times since a wider range of conditions will be reflected in the measurements. Additionally, since NO<sub>x</sub> scavenges ozone near the road, different background ozone concentrations could affect the shape of the NO<sub>x</sub> curves. Although, if this were the case we would expect the NO<sub>2</sub> curve (recall that NO<sub>x</sub> = NO<sub>2</sub> + NO) to exhibit variation as well. Indeed, the results from Singer et al. (2004) and Kim et al. (2004) shown in the NO<sub>x</sub> panel of Figure 8 and plotted as orange triangles, and green inverted triangles, respectively, took measurements using passive sampling over long periods of time. The results from passive sampling of ambient pollution show time averaged concentrations (Palmer 1981) and thus do not reflect short term changes due to traffic or meteorological variation.

Non-alkane hydrocarbons, alkanes, and metal deposition each come from a single study, and show generally good agreement amongst themselves. Results from many different studies, such as UF1 and UF2 particle numbers, show excellent agreement except for elevated values near 400 m for the UF2 particle number. Reponen et al. (2003) reported results included UF2 particle numbers measured at 400 and 800 m from the road only. According to our normalization algorithm, the edge concentration was estimate by a linear fit, which will substantially underpredict the edge value in this case. This is a weakness of the edge normalization procedure as it currently operates. These UF1 particle number data include the previously excluded nighttime study by Zhu et al. (2006) which now integrates more easily with the other data for that pollutant.

Fine particle number, PM<sub>10</sub>, and PM<sub>2.5</sub> again show flat trends. PM<sub>10</sub> no longer shows elevations in the immediate vicinity of the road. This could be due to studies that only reported a single concentration point being excluded from edge normalization. Organic carbon shows an inconclusive trend due to lack of data, and sulfur shows ambiguity, both increasing and decreasing. Clearly, more data is needed on both of these pollutants if patterns are to be discerned.

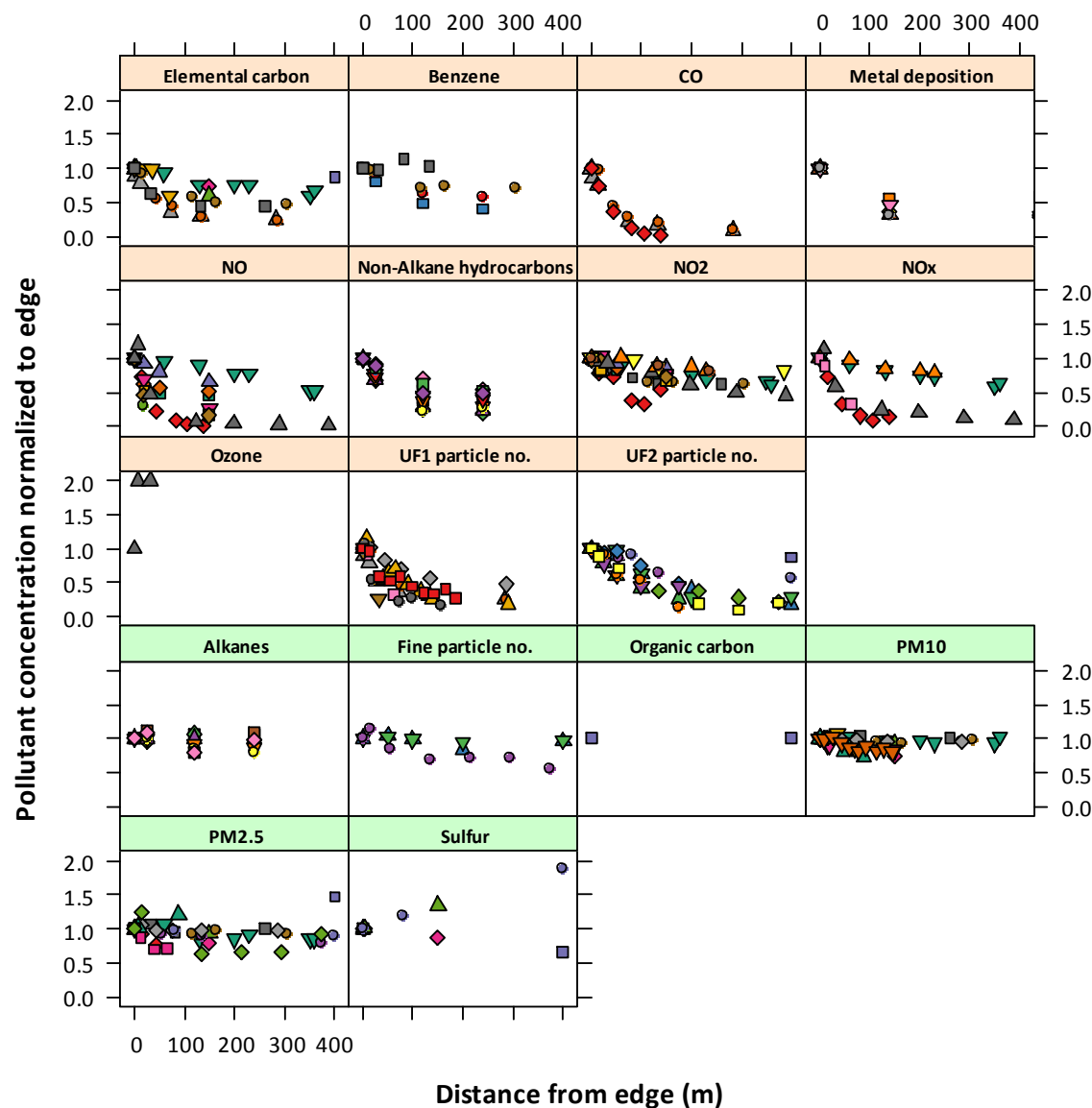


Figure 8: Edge normalized near road air pollution data plotted in separate panels.

Note: Each panel contains all of the data from one pollutant type normalized to the edge of road. The top three rows of panels (orange strip) show those pollutants which vary by distance from road and the bottom two rows (green strip) show those that do not.

The edge normalization procedure appears to perform better than background normalization in terms of grouping the data for any given pollutant more closely together, lending more confidence to the statistical analyses. Future work could extend this analysis to include independent variables which will attempt to quantify the sources of variation in the data. This would allow a further normalization to unit variation of parameters. Some of this work is already being completed by other researchers where complete datasets are available (Zhu et al. 2008).

### 5.2.3.1.2 Analysis of variance

Using methods identical to the background normalization procedure, Table 5 shows the results of the ANOVA conducted on the edge normalized measurements. The table does not include the estimated edge of road values which were all assigned a value of one. Their inclusion led to too many equivalent rankings in the 0 – 80 m bin, and false positives with the Kruskal-Wallis test.

We found that concentrations were significantly different for CO, NO, NO<sub>2</sub>, UF1 and UF2 particle number when comparing the second (80 – 120 m) and third distance bins (> 120 m) to the first (0 – 80 m). NO<sub>x</sub> concentrations were significantly different when comparing the second distance bin to the first. Elemental carbon and surrogates, metal deposition, non-alkane hydrocarbons, and fine particle number showed significant decreases in concentration when comparing the third distance bin to the first.

Table 5 is more densely populated than its background normalized counterpart, owing to an increased amount of data—the degrees of freedom column shows that the majority of pollutants (75 percent) achieved comparisons between all three distance bins compared to 31 percent for background normalized concentrations. Statistically significant relationships, according to the *F* statistic, are approximately equal for both normalizations (11 for edge, 10 for background). Benzene, alkanes, PM<sub>10</sub>, PM<sub>2.5</sub>, and sulfur all show insignificant *F* statistics. In contrast with the results normalized to background concentrations (Table 4), NO and NO<sub>2</sub> concentration changes with road distance are now significant when normalized to the road edge value; they have been moved above the heavy black line to indicate that they vary with distance from road.

The Kruskal-Wallis *p*-value indicates significant differences between groups for all pollutants (*p* < 0.05) except for benzene, ozone, sulfur, and alkanes. Ozone, PM<sub>2.5</sub> and PM<sub>10</sub> show disagreement between tests. For ozone, the difference is due to an inflated Type I error since the *F* statistic shows significance but Kruskal-Wallis does not. This can be caused by violation of any number of the assumptions noted above. The mismatch between test results for PM<sub>2.5</sub> and PM<sub>10</sub> however, results from an increase in Type II error where we failed to reject the null hypothesis with ANOVA, but must for Kruskal-Wallis. This must be due to the non-normality of the distributions of the PM mass concentrations, which are essentially uniform, assuming a flat trend with distance. This result matches the expectation that Kruskal-Wallis has

more power (and is more likely to reject a false null) when distributions are skewed or heavy-tailed. The follow up linear regression does not show significant coefficients for either pollutant, however, and the insignificant mean changes are small. Follow up tests for Kruskal-Wallis to assess the magnitude of between groups variation are possible, but are not performed here. These results indicate, however, that a small trend with distance likely exists for  $PM_{2.5}$  and  $PM_{10}$ . The discussion below uses the mean values given by the regression coefficients, but notes again that these findings should be considered preliminary, pending further analysis of this data regarding deviations from the ANOVA assumptions. Additionally, the means are not directly comparable to the smoother plots, since the loess algorithm uses a weighting function to calculate its predicted values at each data point.

The insignificance of a sulfur trend is likely due to erratic data with a small sample size ( $n = 8$ ). While significant, the trend for fine particle number is the smallest in the table, decreasing by 20 percent from its near road high in the distance bin greater than 120 m. Notwithstanding the disagreement for both measures of PM mass concentrations, the balance of evidence indicated by Table 5 below the heavy black line still shows no or little relationship with distance for these pollutants.

Sample sizes in the middle distance bin are low, indicating that emphasis should be placed on the first and third. Other traffic-related pollutants show significant trends. Both ultrafine particle numbers which decrease in the 80 – 120 m bin, and then decrease further past 120 m. These results show that, on average, concentrations of UF1 particle number decline by 33 percent in the first 80 m, 64 percent by 120 m, and 64 percent beyond 120 m. UF2 particle number declines by 18 percent in the first 80 m, 38 percent by 120 m, and 66 percent past 120 m. CO, NO, and  $NO_x$  show lower normalized concentrations from 80 to 120 m—CO declines by 88 percent, NO by 94 percent, and  $NO_x$  by 87 percent in this distance range. CO and NO show somewhat higher concentrations beyond 120 m, indicating that leveling occurs somewhere in between. Elemental carbon and surrogates, and non-alkane hydrocarbons decay somewhat more slowly, decreasing by 45 percent and 59 percent beyond 120 m from the road. These results attach statistical significance and magnitude to results reported in the literature for concentration gradients associated with air pollutants. The next section presents a comparison between both normalization techniques employed.

Table 5: ANOVA Results for Edge Normalized Pollutant Concentrations

Pollutant	ANOVA		Kruskal-Wallis <i>p</i> -value <sup>a</sup>	0 – 80 m		80 – 120 m		> 120 m	
	Df	<i>F</i>		Coeff. <sup>b</sup>	n	Coeff. <sup>b</sup>	n	Coeff. <sup>b</sup>	n
Benzene	(2,10)	2.89	0.116		4		2		7
CO	(2,13)	9.14**	0.003	0.626	9	-0.538*	2	-0.510**	5
Elemental carbon and surrogates	(2,30)	5.46**	0.01229	0.804	15		1	-0.249**	17
Metal deposition	(1,13)	144.7***	0.00182	1.00	5			-0.615***	10
NO	(2,38)	12.32***	<0.001	0.748	22	-0.691**	2	-0.394***	17
Non-Alkane hydrocarbons	(1,34)	70.5***	<0.001	0.794	12			-0.387***	24
NO <sub>2</sub>	(2,62)	27.7***	<0.001	0.915	25	-0.305***	5	-0.204***	35
NO <sub>x</sub>	(2,22)	3.94*	0.0369	0.737	8	-0.609*	2		15
Ozone	(1,4)	58.4**	0.0603		2				4
UF1 particle no.	(2,41)	14.91***	<0.001	0.670	26	-0.309*	4	-0.309***	14
UF2 particle no.	(2,41)	31.9***	<0.001	0.825	15	-0.208*	7	-0.485***	22
Alkanes	(1,22)	3.71	0.0659		8				16
Fine particle no.	(2,13)	3.88*	0.0184	1.018	4		2	-0.202*	10
PM <sub>10</sub>	(2,45)	0.08	0.0312		24		5		19
PM <sub>2.5</sub>	(2,41)	0.459	0.0485		18		4		22
Sulfur	(2,5)	0.042	0.799		2		1		5

**Note:** The heavy black line separating UF2 particle number from alkanes indicates those pollutants determined to vary with distance (above the line) and those that do not or do so with low magnitude (below the line). Statistical significance is indicated as follows: \*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$ .

Missing values in the table indicate insignificant results as judged by the omnibus  $F$  and the Kruskal-Wallis  $p$ -value, or no data, evidenced by a blank  $n$  for the cell.

<sup>a</sup>The Kruskal-Wallis  $p$ -value was determined using the `kruskal.test()` function in R which performs the non-parametric Kruskal-Wallis test whose null hypothesis is that there is no difference in the mean ranks of the groups.

<sup>b</sup>Refers to the regression coefficients extracted from a linear regression of normalized concentrations on distance bin. The coefficient in the 0 – 80 m bin was the model intercept which represents the mean normalized value in that range, while the coefficients in the other two distance bins represent mean changes relative to the first bin.

#### 5.2.4 Comparison of normalization methods

Both normalization techniques represent the same phenomenon—the atmospheric transport and fate of pollutants as they are carried from the roadway and eventually reach an ambient concentration. This section compares both techniques, making reference to Table 6 and Figures 4 and 7; it relies on the results obtained from local regression due to the statistical difficulties highlighted for ANOVA. The data in Table 6 was taken from above—the percent of near road high at which leveling occurs was entered directly for edge normalization and its reciprocal gives the edge of road elevation above background; the increment above background near the road was entered directly for background normalization and its reciprocal gives the

percent of near road high at which leveling occurs. Table 6 clearly shows the near road pollutants that the meta-analysis determined are high in the vicinity of roads. Summarizing the previously discussed results, the influence of the roadway decays to a background level on average between 80 – 150 m according to edge normalization and by 400 m for the background normalization, depending on the pollutant. Concentrations are elevated from 1.7 – 20 times above background near the road for pollutants with clear trends.

Edge normalization generally shows a more rapid decline to background concentrations than background normalization. This difference arises from different assumptions about background concentrations used by each method. In the background normalization case, we illustrated how a single study could keep the smoothed line elevated above background concentrations simply by reporting a background concentration much lower than the others. In some cases this was due to background concentrations being taken at a location that did not adequately characterize background concentrations in the roadway vicinity, pointing to the need for better matching background concentrations to near road measurements. When normalizing by edge concentrations, these errant background concentrations are not included, and the data yields its own background value by virtue of leveling off. A limitation of the information derived from the edge normalization assessment is the possibility that concentrations level off at different background values which are site-specific, similar to the background normalization. However, to the extent that this is true, we would expect the smoother to be highly variable for these pollutants. If the smoother approaches a horizontal line, it indicates that, generally, consistent background concentrations are measured between studies.

The more gradual reduction to background concentrations suggested by background normalization for certain pollutants (elemental carbon and surrogates, for example) suggests that the processes governing decay to background may be complex or perhaps that there is a selection bias in studies which measure background concentrations. For example, studies focused on communities heavily affected by diesel truck volumes (see, e.g, Lena et al. 2002) would be expected to report above-average roadway increments. The effect of this upward pressure is mitigated in the edge normalized plot, but can still be seen in the elemental carbon curve in Figure 7 which shows increasing concentrations moving from 200 – 400 m, indicating that stability in concentrations has not yet been reached (i.e. the smoother is not an approximately horizontal line). Therefore, while the 80 – 150 m range possibly encompasses the majority of



pollutants for reaching background concentrations, a further distance (up to as much as 400 m), appears to be necessary for other pollutants depending on the site.

**Table 6: Summary of Pollutant Profiles Under Both Normalization Methods**

	Percent of near road high concentration at which leveling begins or background reached		Distance from road at which leveling begins or background reached (m)		Edge of road elevation above background concentration	
	EN	BN <sup>a</sup>	EN	BN	EN <sup>b</sup>	BN
NO	33	--	80	400	3.0	--
Elemental carbon and surrogates	60	56	90	400	1.7	1.8
NO <sub>x</sub>	54	59	100	400	1.9	1.7
CO	10	5	120	330	10	20
Non-Alkane Hydrocarbons	35	45	150	150	2.9	2.2
UF1 particle no.	30	25	150	200	3.3	4
Metal deposition	--	33	--	175	--	3
Benzene	70	33	300	75	1.4	3
NO <sub>2</sub>	67	--	350	--	1.5	--
UF2 particle no.	--	20	--	--	--	5

**Note:** EN=Edge normalization; BN=Background normalization. Benzene, NO<sub>2</sub>, and UF2 particle number (below the heavy line in the table) show continuous declines in concentration and no leveling over the entire distance range, depending on normalization method. Table entries are sorted based on the edge normalized distance at which background concentrations are reached. The results presented in this table are not directly comparable to the regression results presented alongside the ANOVA in Tables 4 and 5, since the smoother uses a weighting function to calculate its estimations, as opposed to a straight mean in the linear regression. --Data unavailable, or pollutant showed continuous declines.

<sup>a</sup>Calculated as the inverse of the edge of road elevation above background concentrations.

<sup>b</sup>Calculated as the inverse of the percent of near road high at which leveling occurs.

The trends indicated by both methods are still broadly consistent, not considering the distance at which background is reached. NO and NO<sub>2</sub> are the only pollutants that do not have consistent behavior, but the sample sizes for background normalization used in Figure 7 for NO (n = 5) and NO<sub>2</sub> (n = 9) are lower than that for the edge by an order of magnitude (n = 51 and n = 80, respectively) so it is likely that the edge normalization is more reliable. Similarities are also shown when comparing the last two columns of Table 6 which indicate a general agreement in terms of the increment at the roadway above background concentrations (at least to an order of magnitude, in the case of CO, and much closer for elemental carbon and surrogates, NO<sub>x</sub>, and UF1 particle number). This is likely because there is simply more data in this range, so it is unlikely that any single measurement will affect the results. The agreement between methods in the vicinity of the roadway was still contingent on the exclusion of some of the data from the

background normalized smoothers. This speaks to the strength of the edge normalization technique.

## 6 LIMITATIONS AND FUTURE WORK

Several limitations to generalizability can be noted. Most importantly, the data contained in this meta-analysis was aggregated from studies with non-uniform sampling procedures and non-uniform locations (i.e. different roadway, geographic, and meteorological conditions). Additionally, the sampling periods employed by each investigator varied between several hours and several days. It is likely that concentrations averaged over longer time periods will vary less than those measured during shorter time intervals. Background concentrations were not always averaged over the same period as the associated near road measurements; this would tend to overstate the increment associated with the roadway if background represented generally longer times. Normalization will not necessarily remove these differences since they may imply different atmospheric processes (as seen in the data for NO and NO<sub>x</sub> normalized to the edge concentrations). Depending on the extent of these deviations, the reported results may be called into question. However, the results as reported follow accepted knowledge on near roadway pollutants. The behavior of the smoothers also reflects the strength of the normalization procedures.

These and other issues should be addressed in future work. Additional analyses could which will model the observed near roadway concentrations as functions of a number of independent variables using ordinary least squares regression. The existing literature (see studies in Appendix B) showed that several factors consistently influence observed concentrations—wind speed and direction, other meteorology, traffic volumes and fleet composition—future work could use the dataset assembled for this work to further assess the magnitude and relative contribution of each factor to observed concentrations. However, such assessments may be difficult because of inconsistency among the factors reported within each study and the problems of acquisition of secondary data. Rijnders et al. (2001), Janssen et al. (2001), and Ntziachristos et al. (2007b) performed similar assessments using complete datasets from individual sites to attribute variations in observed pollution concentrations to changes in traffic volumes or meteorological characteristics.

In addition, a closer analysis of the statistical properties of the data should be undertaken. This could be as simple as assessing the performance of each of the nonlinear models used for edge normalization. If they perform well, they could be used to simulate a large amount of data which could be used to further test the assumptions of the statistics performed. The simulated data could be used in a theoretical analysis of the suitability of near road data to parametric and nonparametric analyses and would work to highlight the differences observed between the ANOVA and Kruskal-Wallis test.

An obvious extension to this work on observed concentrations is to apply the same methods to the public health literature that examines health outcomes in relation to distance from road or exposure to differences in traffic volumes or composition. Such work will need to take into consideration the risk-weighted importance of observed concentrations. For example, it is likely that chronic exposure to diesel exhaust is carcinogenic (EPA 2002), so the twice elevated concentrations of elemental carbon (which are a diesel exhaust marker) observed near the road (Table 6) might be much more worrisome from a public health perspective than the 10 – 20 times elevated concentrations of CO. Overlaying similar statistical analysis from public health with the results presented here will help to identify causation and to prioritize mitigation options and policymaking.

The only data excluded during this assessment were data collected in a substantially different time period (1970s vs. 1990s or 2000s) or under substantially different meteorological conditions (nighttime vs. daytime) than the majority of the data. However, we also showed that there were likely issues associated with averaging time (e.g., days or weeks vs. longer-term measurements of a year or more), and this affected the observed trends. Collecting longer-term measurements would improve understanding. Also, investigating criteria by which more data could reasonably be excluded or grouped further should strengthen the results.

There is a need for a closer investigation of PM data disaggregated by chemical species (nitrates, sulfates, ammonium). Presently, these data are not included in our database. Differences according to size fraction measured and distance from roadway are fruitful areas for further work especially since the health implications of PM are known to vary according to composition (Magari et al. 2002). Ntziachristos et al. (2007a) also called for increased classification of near road PM composition. Additionally, particle size distributions were measured by many of the studies, but we have not discussed them here. These measurements

have important implications for source apportionment and will likely play a key role in further analyses.

## 7 CONCLUSIONS

This work completed a meta-analysis of observed near roadway air pollution concentrations. A database was used to store air pollution concentration data from near roadway studies. We showed that the raw data was too disparate to facilitate meaningful comparison without normalization, and employed two different normalization methods to improve data assessment.

Background normalization was shown to perform inconsistently, because of differences in what constitutes background measurements from site to site. Background concentrations in future studies must be more closely matched to the area near the road. Normalizing to the edge of road concentration, as estimated from either linear or nonlinear models (depending on the set of measurements), performed better than background normalization in two respects. First, it grouped the data for individual pollutants more closely together, lending more confidence to subsequent statistical analyses even though important features of the data, such as averaging time, were not considered. Second, it did not rely on the problematic measurements which tended to show a longer decay to background concentrations. These techniques can contribute to a more comprehensive modeling of near roadway concentrations using the data available in the literature.

The results of applying a local regression to normalized data showed that some pollutants are significantly elevated above background concentrations near roadways. NO, NO<sub>x</sub>, CO, elemental carbon and surrogates, metal deposition, non-alkane hydrocarbons, and ultrafine particle numbers were all elevated in the near road region from 1.7 – 20 times above background values, combining the results from both normalization methods. In spite of their differences, both methods showed broad agreement in the distance-from-road range above which concentrations at the roadway edge are elevated, owing to the large amount of data within the first 50 m.

According to the edge normalization, NO, NO<sub>x</sub>, CO, elemental carbon and surrogates, metal deposition, non-alkane hydrocarbons, and UF1 particle numbers decreased to background between 80 and 150 m from the edge of road. Background normalized results did not converge on background until 400 m, but this likely overstates the distance, because of the shortcomings

with this method—most importantly that the definition of what constitutes a background concentration is not consistent between studies. Parametric (ANOVA) and nonparametric (the Kruskal-Wallis test, to account for violation of ANOVA assumptions) statistical analyses were performed, and found statistically significant relationships with distance from road for NO, NO<sub>x</sub>, CO, elemental carbon and surrogates, metal deposition, non-alkane hydrocarbons, and ultrafine particle numbers. Benzene and NO<sub>2</sub> showed flatter trends. These pollutants reached background levels by 300 – 350 m, but the leveling was not as apparent as for the others.

A final group of pollutants showed insignificant or very small but significant trends with distance including PM<sub>10</sub>, PM<sub>2.5</sub>, sulfur, alkanes, and fine particle number (here defined as particles with aerodynamic diameter greater than 300 nm). In areas with relatively high regional background concentrations of PM<sub>10</sub> and PM<sub>2.5</sub>, the incremental contribution from the road is small in comparison, rendering any roadway contribution marginal. The PM mass concentration data available for this study showed significant trends according to the Kruskal-Wallis test but not with ANOVA, which likely results from the nonparametric test reducing Type II error because of the relatively flat distribution of PM<sub>10</sub> and PM<sub>2.5</sub> concentrations near roads. Further work is needed on the underlying assumptions associated with ANOVA, but the generally good agreement of the Kruskal-Wallis and ANOVA results lends support to the use of either method for the detection of magnitude difference by distance groups in near road air pollution. After the test for differences between means, a linear regression was performed for concentration on the distance groupings, yielding the average magnitudes of concentration declines in three distance groups. This work also assessed the suitability of near road pollution analyses to employ ANOVA and local regression. These methods performed well in terms of generating intuitive results and confirmed the use of the normalization techniques employed.

Finally, further work in this area could focus on mitigation. Mitigation has only been sparsely discussed, except by EPA (Bowker et al. 2007; Baldauf et al. 2008a). Our analyses can be used to identify where the largest gains can be made by mitigation strategies directed at certain pollutants. In general, efforts directed further than 150 m from the road are likely to produce diminished results compared to those located closer to the edge.

## REFERENCES

- Ashbaugh, L. L., R. G. Flocchini, D. Chang, V. Garza, O. F. Carvacho, T. A. James and R. T. Matsumura. (1996). Traffic Generated PM<sub>10</sub> "Hot Spots". University of California, Davis. C. N. L. Air Quality Group, Davis, CA.
- Ashbaugh, L. L., R. G. Flocchini, R. T. Matsumura, T. A. James, O. F. Carvacho, C. Tsubamoto and M. Brown. (1998). Wintertime Traffic Generated PM<sub>10</sub> Hot Spots. University of California, Davis. C. N. L. Air Quality Group, Davis, CA. Contract number 43X878.
- Bae, C. H. C., G. Sandlin, A. Bassok and S. Kim (2007). "The exposure of disadvantaged populations in freeway air-pollution sheds: a case study of the Seattle and Portland regions." Environment and Planning B: Planning and Design **34**: 154-170.
- Baldauf, R., E. Thoma, A. Khlystov, V. Isakov, G. Bowker, T. Long and R. Snow (2008a). "Impacts of noise barriers on near-road air quality." Atmospheric Environment **42**(32): 7502-7507.
- Baldauf, R. W., G. E. Bowker, A. Khlystov, D. Heist, S. Perry, V. Isakov, L. Brixley, E. Thoma, T. Long and G. Hagler (2008b). The Influence of Roadway Configuration on Near Road Air Quality. Air & Waste Management Association 101st Annual Conference and Exhibition. Oregon Convention Center, Portland, Oregon.
- Behrens, T., D. Taeger, W. Maziak, H. Duhme, P. Rzehak, S. K. Weiland and U. Keil (2004). "Self-reported traffic density and atopic disease in children. Results of the ISAAC Phase III survey in Muenster, Germany." Pediatric Allergy and Immunology **15**(4): 331-339.
- Bowker, G. E., R. Baldauf, V. Isakov, A. Khlystov and W. Petersen (2007). "The effects of roadside structures on the transport and dispersion of ultrafine particles from highways." Atmospheric Environment **41**(37): 8128-8139.
- Broderick, B. M. and R. T. O'Donoghue (2007). "Spatial variation of roadside C2-C6 hydrocarbon concentrations during low wind speeds: Validation of CALINE4 and COPERT III modelling." Transportation Research Part D: Transport and Environment **12**(8): 537-547.
- Brugge, D., J. Durant and C. Rioux (2007). "Near-highway pollutants in motor vehicle exhaust: A review of epidemiologic evidence of cardiac and pulmonary health risks." Environmental Health **6**(1): 23.
- Brunekreef, B. and S. T. Holgate (2002). "Air pollution and health." Lancet **360**: 1233 - 1242.
- Brunekreef, B., N. A. Janssen, J. de Hartog, H. Harssema, M. Knape and P. van Vliet (1997). "Air pollution from truck traffic and lung function in children living near motorways." Epidemiology **8**: 298-303.

- Cahill, T., P. Feeney, R. Flocchini and T. Dunn. (1974). Particulate Matter from California Freeways. California Air Resources Board. Sacramento, CA.
- Cahill, T., D. Sperling, D. Chang, E. Gearhart, O. Carvacho and L. Ashbaugh. (1994). PM<sub>10</sub> "Hot Spot" Emissions from California Roads. University of California, Davis. A. Q. Group, Davis, CA.
- Caltrans. (2008). "Highway Design Manual." Retrieved November 28, 2008, from <http://www.dot.ca.gov/hq/oppd/hdm/hdmtoc.htm>.
- Cannon, H. L. and J. M. Bowles (1962). "Contamination of Vegetation by Tetraethyl Lead." Science **137**(3532): 765-766.
- CARB. (2005). Air Quality and Land Use Handbook: A Community Health Perspective. California Air Resources Board. Sacramento, CA.
- Chan, C.-C., H. Ozkaynak, J. Spengler and L. Sheldon (1991). "Driver exposure to volatile organic compounds, CO, ozone, and NO<sub>2</sub> under different driving conditions." Environmental Science & Technology **25**: 964-972.
- Clements, A., Y. Jia, M. P. Fraser, Y. Zhu, J. Pudota, A. DenBleyker, E. Michel, D. R. Collins, E. McDonald-Buller and D. T. Allen (2008). Air Pollutant Concentrations near Texas Roadways: Chemical Transformations of Pollutants. Air & Waste Management Association Annual Conference and Exhibition. Oregon Convention Center, Portland, Oregon.
- Cleveland, W. S. (1979). "Robust Locally Weighted Regression and Smoothing Scatterplots." Journal of the American Statistical Association **74**(368): 829-836.
- Cleveland, W. S. and S. J. Devlin (1988). "Locally Weighted Regression: An Approach to Regression Analysis by Local Fitting." Journal of the American Statistical Association **83**(403): 596-610.
- Cohen, J., R. Cook, C. R. Bailey and E. Carr (2005). "Relationship between motor vehicle emissions of hazardous pollutants, roadway proximity, and ambient concentrations in Portland, Oregon." Environmental Modelling & Software **20**: 7-12.
- Cyrys, J., J. Heinrich, G. Hoek, K. Meliefste, M. Lewne, U. Gehring, T. Bellander, P. Fischer, P. Van Vliet, M. Brauer, H. E. Wichmann and B. Brunekreef (2003). "Comparison between different traffic-related particle indicators: Elemental carbon (EC), PM<sub>2.5</sub> mass, and absorbance." Journal of Exposure Analysis and Environmental Epidemiology **13**: 134-143.
- Dockery, D. W., C. A. Pope, X. Xu, J. D. Spengler, J. H. Ware, M. E. Fay, B. G. Ferris and F. E. Speizer (1993). "An association between air pollution and mortality in six U.S. cities." New England Journal of Medicine **329**: 1753 - 9.

- Donaldson, K., V. Stone, A. Clouter, L. Renwick and W. MacNee (2001). "Ultrafine particles." Occupational and Environmental Medicine **58**(3): 211-216.
- Duhme, H., S. Weiland, U. Keil, B. Kraemer, M. Schmid, M. Stender and L. Chambless (1996). "The association between self-reported symptoms of asthma and allergic rhinitis and self-reported traffic density on street residence in adolescents." Epidemiology **7**: 578-582.
- Edwards, J., S. Walters and R. K. Griffiths (1994). "Hospital admissions for asthma in preschool children: relationship to major roads in Birmingham, United Kingdom." Archives of Environmental Health **49**(4): 223-227.
- EPA. (2002). Health Assessment Document for Diesel Engine Exhaust. Washington, DC. EPA/600/8-90/057F.
- Fischer, P. H., G. Hoek, H. van Reeuwijk, D. J. Briggs, E. Lebret, J. H. van Wijnen, S. Kingham and P. E. Elliott (2000). "Traffic-related differences in outdoor and indoor concentrations of particles and volatile organic compounds in Amsterdam." Atmospheric Environment **34**(22): 3713-3722.
- Gauderman, W. J., E. Avol, F. Lurmann, N. Kunzli, F. Gilliland, J. Peters and R. McConnell (2005). "Childhood asthma and exposure to traffic and nitrogen dioxide." Epidemiology **16**: 737-743.
- Gidhagen, L., C. Johansson, G. Omstedt, J. Langner and G. Olivares (2004). "Model simulations of NO<sub>x</sub> and ultrafine particles close to a Swedish highway." Environmental Science & Technology **38**(24): 6730-6740.
- Gilbert, N. L., M. S. Goldberg, B. Beckerman, J. R. Brook and M. Jerrett (2005). "Assessing spatial variability of ambient nitrogen dioxide in Montreal, Canada, with a land-use regression model." Journal of the Air & Waste Management Association **55**(8): 1059-1063.
- Gilbert, N. L., M. S. Goldberg, J. R. Brook and M. Jerrett (2007). "The influence of highway traffic on ambient nitrogen dioxide concentrations beyond the immediate vicinity of highways." Atmospheric Environment **41**(12): 2670-2673.
- Gilbert, N. L., S. Woodhouse, D. M. Stieb and J. R. Brook (2003). "Ambient nitrogen dioxide and distance from a major highway." Science of the Total Environment **312**(1-3): 43-46.
- Gilbert, R. O. (1987). Statistical Methods for Environmental Pollution Monitoring. New York, NY, Van Nostrand Reinhold Co.
- Glass, G. V. (1976). "Primary, Secondary, and Meta-Analysis of Research." Educational Researcher **5**(10): 3-8.
- Gramotnev, G. and Z. Ristovski (2004). "Experimental investigation of ultra-fine particle size distribution near a busy road." Atmospheric Environment **38**: 1767-1776.



- Gunier, R. B., A. Hertz, J. von Behren and P. Reynolds (2003). "Traffic density in California: Socioeconomic and ethnic differences among potentially exposed children." Journal of Exposure Analysis and Environmental Epidemiology **13**(3): 240-246.
- Hagler, G., E. D. Thoma, R. Baldauf, T. R. Long and R. F. Snow (2008). Impact of Meteorology, Traffic Characteristics, and Distance from Roadway on Roadside Concentrations of Ultrafine Particulate Matter. Air & Waste Management Association 101st Conference and Exhibition. Oregon Convention Center, Portland, Oregon.
- Harrison, R. M., R. Tilling, M. S. C. Romero, S. Harrad and K. Jarvis (2003). "A study of trace metals and polycyclic aromatic hydrocarbons in the roadside environment." Atmospheric Environment **37**: 2391-2402.
- Hitchins, J., L. Morawska, R. Wolff and D. Gilbert (2000). "Concentrations of submicrometre particles from vehicle emissions near a major road." Atmospheric Environment **34**: 51-59.
- Hoek, G., K. Meliefste, J. Cyrys, M. Lewne, T. Bellander, M. Brauer, P. Fischer, U. Gehring, J. Heinrich, P. van Vliet and B. Brunekreef (2002). "Spatial variability of fine particle concentrations in three European areas." Atmospheric Environment **36**: 4077-4088.
- Holmes, N. S. and L. Morawska (2006). "A review of dispersion modelling and its application to the dispersion of particles: An overview of different dispersion models available." Atmospheric Environment **40**(30): 5902-5928.
- Houston, D., P. Ong, J. Wu and A. Winer (2006). "Proximity of licensed childcare to concentrated vehicle pollutants near major roadways." American Journal of Public Health **96**(9): 1611-1617.
- Houston, D., J. Wu, P. Ong and W. A. (2004). "Structural disparities of urban traffic in Southern California: implications for vehicle-related air pollution exposure in minority and high-poverty neighborhoods." Journal of Urban Affairs **26**: 565-592.
- Janssen, N. A. H., P. H. N. van Vliet, F. Aarts, H. Harssema and B. Brunekreef (2001). "Assessment of exposure to traffic related air pollution of children attending schools near motorways." Atmospheric Environment **35**: 3875-3884.
- Janssen, N. A. H., D. F. M. VanMansom, K. VanDerJagt, H. Harssema and G. Hoek (1997). "Mass concentration and elemental composition of airborne particulate matter at street and background locations." Atmospheric Environment **31**(8): 1185-1193.
- Karner, A., D. Eisinger, S. Bai and D. Niemeier (2008). Mitigating Diesel Truck Impacts in Environmental Justice Communities: Transportation Planning and Air Quality in Barrio Logan, San Diego. Paper Presented at the 88th Annual Meeting of the Transportation Research Board. Washington, DC.

- Kim, J. J., S. Smorodinsky, M. Lipsett, B. C. Singer, A. T. Hodgson and B. Ostro (2004). "Traffic-related air pollution near busy roads - The East Bay children's respiratory health study." American Journal of Respiratory and Critical Care Medicine **170**: 520-526.
- Kuhn, T., S. Biswas, P. Fine, M. Geller and C. Sioutas (2005a). "Physical and Chemical Characteristics and Volatility of PM in the Proximity of a Light-Duty Vehicle Freeway." Aerosol Science & Technology **39**(4): 347-357.
- Kuhn, T., S. Biswas and C. Sioutas (2005b). "Diurnal and seasonal characteristics of particle volatility and chemical composition in the vicinity of a light-duty vehicle freeway." Atmospheric Environment **39**(37): 7154-7166.
- Kwon, J. (2005). Development of a RIOPA database and evaluation of the effect of proximity on the potential residential exposure to VOCs from ambient sources. New Brunswick, NJ, Rutgers, The State University of New Jersey. **Ph.D Thesis**.
- Lagerwerff, J. V. and A. W. Specht (1970). "Contamination of roadside soil and vegetation with cadmium, nickel, lead, and zinc." Environmental Science & Technology **4**(7): 583-586.
- Lena, T. S., V. Ochieng, M. Carter, J. Holguin-Veras and P. L. Kinney (2002). "Elemental carbon and PM<sub>2.5</sub> levels in an urban community heavily impacted by truck traffic." Environmental Health Perspectives **110**: 1009-1015.
- Levy, J. I., D. H. Bennett, S. J. Melly and J. D. Spengler (2003). "Influence of traffic patterns on particulate matter and polycyclic aromatic hydrocarbon concentrations in Roxbury, Massachusetts." Journal of Exposure Analysis and Environmental Epidemiology **13**(5): 364-371.
- Levy, J. I., J. K. Hammitt and J. D. Spengler (2000). "Estimating the mortality impacts of particulate matter: what can be learned from between-study variability?" Environmental Health Perspectives **108**(2): 109-17.
- Lin, S., P. M. Munsie, S. Hwang, E. Fitzgerald and M. R. Cayo (2002). "Childhood asthma hospitalization and residential exposure to state route traffic." Environmental Research, Section A **88**: 73-81.
- Luginaah, I., X. Xu, K. Y. Fung, A. Grgicak-Mannion, J. Wintermute, A. Wheeler and J. Brook (2006). "Establishing the Spatial Variability of Ambient Nitrogen Dioxide in Windsor, Ontario." International journal of Environmental Studies **63**(4): 487-500.
- Magari, S. R., J. Schwartz, P. L. Williams, R. Hauser, T. J. Smith and D. C. Christiani (2002). "The association of particulate air metal concentrations with heart rate variability." Environmental Health Perspectives **110**(9): 875-80.
- McConnell, R., K. Berhane, L. Yao, M. Jerrett, F. W. Lurmann, F. Gilliland, N. Kunzli, W. J. Gauderman, E. Avol and J. Peters (2006). "Traffic, susceptibility, and childhood asthma." Environmental Health Perspectives **114**: 766-772.

- Molnar, P., S. Janhall and M. Hallquist (2002). "Roadside measurements of fine and ultrafine particles at a major road north of Gothenburg." Atmospheric Environment **36**: 4115-4123.
- Morawska, L., S. Thomas, D. Gilbert, C. Greenaway and E. Rijnders (1999). "A study of the horizontal and vertical profile of submicrometer particles in relation to a busy road." Atmospheric Environment **33**(8): 1261-1274.
- Mudd, G. M. and M. Diesendorf (2008). "Sustainability of Uranium Mining and Milling: Toward Quantifying Resources and Eco-Efficiency." Environmental Science & Technology **42**(7): 2624-2630.
- Murakami, M., M. Ono and K. Tamura (1990). "Health problems of residents along heavy-traffic roads." Journal of Human Ergology **19**: 101-106.
- Myers, J. L. and A. D. Well (1991). Research Design and Statistical Analysis. New York, NY, HarperCollins.
- Nanzetta, M. K. and B. A. Holmen (2004). "Roadside particle number, distributions and relationships between number concentrations, meteorology, and traffic along a northern California freeway." Journal of the Air & Waste Management Association **54**: 540-554.
- Nel, A. (2005). "Air Pollution-Related Illness: Effects of Particles." Science **308**(5723): 804-806.
- Nitta, H., T. Sato, S. Nakai, K. Maeda, S. Aoki and M. Ono (1993). "Respiratory health associated with exposure to automobile exhaust.1. Results of cross-sectional studies in 1979, 1982, and 1983." Archives of Environmental Health **48**(1): 53-58.
- Ntziachristos, L., Z. Ning, M. D. Geller, R. J. Sheesley, J. J. Schauer and C. Sioutas (2007a). "Fine, ultrafine and nanoparticle trace element compositions near a major freeway with a high heavy-duty diesel fraction." Atmospheric Environment **41**(27): 5684-5696.
- Ntziachristos, L., Z. Ning, M. D. Geller and C. Sioutas (2007b). "Particle concentration and characteristics near a major freeway with heavy-duty diesel traffic." Environmental Science & Technology **41**(7): 2223-2230.
- O'Donoghue, R. T. and B. M. Broderick (2007). "Spatial variation of roadside C2-C6 hydrocarbon concentrations during low wind speeds: A note." Transportation Research Part D: Transport and Environment **12**(8): 589-595.
- Palmes, E. D. (1981). "Development and application of a diffusional sampler for NO<sub>2</sub>." Environment International **5**(2): 97-100.
- Pedersen, C. B., O. Raaschou-Nielsen, O. Hertel and P. B. Mortensen (2004). "New Directions: Air pollution from traffic and schizophrenia risk." Atmospheric Environment **38**(22): 3733-3734.

- Peters, A., S. von Klot, M. Heier, I. Trentinaglia, A. Hormann, H. E. Wichmann and H. Lowel (2004). "Exposure to traffic and the onset of myocardial infarction." The New England Journal of Medicine **351**: 1861 - 70.
- Reponen, T., S. A. Grinshpun, S. Trakumas, D. Martuzevicius, Z. M. Wang, G. LeMasters, J. E. Lockey and P. Biswas (2003). "Concentration gradient patterns of aerosol particles near interstate highways in the Greater Cincinnati airshed." Journal of Environmental Monitoring **5**: 557-562.
- Reynolds, P., J. Von Behren, R. B. Gunier, D. E. Goldberg and A. Hertz (2004). "Residential exposure to traffic in California and childhood cancer." Epidemiology **15**(1): 6-12.
- Rijnders, E., N. A. H. Janssen, P. H. N. van Vliet and B. Brunekreef (2001). "Personal and outdoor nitrogen dioxide concentrations in relation to degree of urbanization and traffic density." Environmental Health Perspectives **109**: 411-417.
- Rodes, C. E. and D. M. Holland (1981). "Variations of NO, NO<sub>2</sub> and O<sub>3</sub> concentrations downwind of a Los Angeles freeway." Atmospheric Environment **15**(3): 243-250.
- Roemer, W. H. and J. H. van Wijnen (2001). "Differences among black smoke, PM<sub>10</sub>, and PM<sub>1.0</sub> levels at urban measurement sites." Environmental Health Perspectives **109**: 151-154.
- Roorda-Knappe, M. C., N. A. H. Janssen, J. J. De Hartog, P. H. N. Van Vliet, H. Harssema and B. Brunekreef (1998). "Air pollution from traffic in city districts near major motorways." Atmospheric Environment **32**(11): 1921-1930.
- Ross, Z., M. Jerrett, K. Ito, B. Tempalski and G. D. Thurston (2007). "A land use regression for predicting fine particulate matter concentrations in the New York City region." Atmospheric Environment **41**(11): 2255-2269.
- Ruellan, S. and H. Cachier (2001). "Characterisation of fresh particulate vehicular exhausts near a Paris high flow road." Atmospheric Environment **35**: 453-468.
- Sabin, L. D., J. H. Lim, M. T. Venezia, A. M. Winer, K. C. Schiff and K. D. Stolzenbach (2006). "Dry deposition and resuspension of particle-associated metals near a freeway in Los Angeles." Atmospheric Environment **40**: 7528-7538.
- Sapkota, A. and T. J. Buckley (2003). "The mobile source effect on curbside 1,3-butadiene, benzene, and particle-bound polycyclic aromatic hydrocarbons assessed at a tollbooth." Journal of the Air & Waste Management Association **53**: 740-748.
- Savitz, D. A. and L. Feingold (1989). "Association of childhood cancer with residential traffic density." Scandinavian Journal of Work, Environment & Health **15**(5): 360-363.
- Schwartz, J. (1994). "Air Pollution and Daily Mortality: A Review and Meta Analysis." Environmental Research **64**(1): 36-52.

- Sexton, K. and P. B. Ryan (1988). Assessment of human exposure to air pollution: methods, measurements and models. Air Pollution, the Automobile and Public Health. A. Y. Watson, R. R. Bates and D. Kennedy. New York, NY, National Academy Press: 207-238.
- Shi, J. P., A. A. Khan and R. M. Harrison (1999). "Measurements of ultrafine particle concentration and size distribution in the urban atmosphere." Science of the Total Environment **235**: 51 - 64.
- Singer, B. C., A. T. Hodgson, H. Toshifumi and J. J. Kim (2004). "Passive measurement of nitrogen oxides to assess traffic-related pollutant exposure for the East Bay Children's Respiratory Health Study." Atmospheric Environment **38**: 393-403.
- Smargiassi, A., M. Baldwin, C. Pilger, R. Dugandzic and M. Brauer (2005). "Small-scale spatial variability of particle concentrations and traffic levels in Montreal: a pilot study." Science of the Total Environment **338**(3): 243-251.
- Stieb, D. M., S. Judek and R. T. Burnett (2002). "Meta-Analysis of Time-Series Studies of Air Pollution and Mortality: Effects of Gases and Particles and the Influence of Cause of Death, Age, and Season." Journal of the Air & Waste Management Association **52**: 470-484.
- Tiitta, P., T. Raunemaa, J. Tissari, T. Yli-Tuomi, A. Leskinen, J. Kukkonen, J. Harkonen and A. Karppinen (2002). "Measurements and modelling of PM<sub>2.5</sub> concentrations near a major road in Kuopio, Finland." Atmospheric Environment **36**: 4057-4068.
- van Vliet, P., M. Knape, J. de Hartog, M. Janssen, H. Harssema and B. Brunekreef (1997). "Motor vehicle exhaust and chronic respiratory symptoms in children living near freeways." Environmental Research **74**: 122-132.
- Vardoulakis, S., N. Gonzalez-Flesca, B. E. A. Fisher and K. Pericleous (2005). "Spatial variability of air pollution in the vicinity of a permanent monitoring station in central Paris." Atmospheric Environment **39**: 2725-2736.
- Venables, W. N. and B. D. Ripley (2002). Modern Applied Statistics with S, Fourth Edition. New York, NY, Springer-Verlag.
- Weijers, E. P., A. Y. Khlystov, G. P. A. Kos and J. W. Erisman (2004). "Variability of particulate matter concentrations along roads and motorways determined by a moving measurement unit." Atmospheric Environment **38**: 2993-3002.
- Weiland, S. K., K. A. Mundt, A. Ruckman and U. Keil (1994). "Self-reported wheezing and allergic rhinitis in children and traffic density on street of residence." Annals of Epidemiology **4**: 243-247.
- Westerdahl, D., S. Fruin, T. Sax, P. M. Fine and C. Sioutas (2005). "Mobile platform measurements of ultrafine particles and associated pollutant concentrations on freeways and residential streets in Los Angeles." Atmospheric Environment **39**(20): 3597-3610.

- Wjst, M., P. Reitmeir, S. Dodd, A. Wulff, T. Nicolai, E. F. von Loeffelholz-Colberg and E. von Mutius (1993). "Road traffic and adverse effects on respiratory health in children." British Medical Journal **307**: 596 - 307.
- Wu, Y., J. M. Hao, L. X. Fu, J. N. Hu, Z. S. Wang and U. Tang (2003). "Chemical characteristics of airborne particulate matter near major roads and at background locations in Macao, China." Science of the Total Environment **317**: 159-172.
- Wyler, C., C. Braun-Fahrlander, N. Kunzli, C. Schindler, U. Ackermann-Liebrich, A. P. Perruchoud, P. Leuenberger and B. Wuthrich (2000). "Exposure to Motor Vehicle Traffic and Allergic Sensitization." Epidemiology **11**(4): 450-456.
- Zhang, K. M., A. S. Wexler, Y. F. Zhu, W. C. Hinds and C. Sioutas (2004). "Evolution of particle number distribution near roadways. Part II: the 'Road-to-Ambient' process." Atmospheric Environment **38**: 6655-6665.
- Zhou, Y. and J. Levy (2007). "Factors influencing the spatial extent of mobile source air pollution impacts: a meta-analysis." BMC Public Health **7**(1): 89.
- Zhu, Y., A. Eiguren-Fernandez, W. C. Hinds and A. H. Miguel (2007). "In-cabin commuter exposure to ultrafine particles on Los Angeles freeways." Environmental Science & Technology **41**(7): 2138-2145.
- Zhu, Y., W. C. Hinds, S. Kim, S. Shen and C. Sioutas (2002a). "Study of ultrafine particles near a major highway with heavy-duty diesel traffic." Atmospheric Environment **36**(27): 4323-4335.
- Zhu, Y., W. C. Hinds, S. Kim and C. Sioutas (2002b). "Concentration and size distribution of ultrafine particles near a major highway." Journal of the Air and Waste Management Association **52**(9): 1032 - 1042.
- Zhu, Y., T. Kuhn, P. Mayo and W. C. Hinds (2006). "Comparison of daytime and nighttime concentration profiles and size distributions of ultrafine particles near a major highway." Environmental Science & Technology **40**(8): 2531-2536.
- Zhu, Y. F., W. C. Hinds, S. Shen and C. Sioutas (2004). "Seasonal trends of concentration and size distribution of ultrafine particles near major highways in Los Angeles." Aerosol Science and Technology **38**(S1): 5-13.
- Zhu, Y. F., J. Pudota, A. DenBleyker, E. Michel, E. McDonald-Buller, M. P. Fraser, D. Collins, Y. Jia, A. Clements and D. Allen (2008). Physical Transformation of Ultrafine Particles near Three Texas Roadways. Air & Waste Management Association 101st Annual Conference and Exhibition. Oregon Convention Center, Portland, Oregon.
- Zmirou, D., S. Gauvin, I. Pin, I. Momas, F. Sahraoui, J. Just, Y. Le Moullec, F. Bremont, S. Cassadou, P. Reungoat, M. Albertini, N. Lauvergne, M. Chiron and A. Labbe (2004). "Traffic related air pollution and incidence of childhood asthma: results of the Vesta case-control study." Journal of Epidemiology and Community Health **58**(1): 18-23.

## Appendix A: R code

The code contained in this appendix can be used to reproduce all of the plots contained in this report when combined with the Microsoft Access dataset. All code was run using R 2.7.2 on Windows XP and utilized Tinn-R—a GUI/editor for R. Scripts were separated according to logical groupings, and a comment (denoted in R using a pound “#” sign) noting a new filename below indicates the beginning of a new script.

```
#####
## Filename: Header.r
## Script purpose: Load required libraries and define functions
#####

## Required libraries
require(stats)
require(splines)
require(akima)
require(MASS)
library(locfit)

## Font definitions
windowsFonts(Cambria = windowsFont("TT Cambria Bold"))
windowsFonts(Calibri = windowsFont("TT Calibri"))

## Miscellaneous functions
countNonZero <- function (vector1) {
  vector1[is.na(vector1)] <- 0;
  count <- vector1 > 0;
  sum(count)
}

negexp <- selfStart(model = ~b0 + b1*exp(-x/th),
initial=negexp.SSival,parameters=c("b0","b1","th"),template =
function(x,b0,b1,th) {} )
edgeConcExp <- function(Conc, Dis) {
  ## If either of the concentration valid numbers or distance valid numbers
  are less than 2, then connect the dots
  if(countNonZero(Conc) <= 2 | countNonZero(Dis) <= 2) {
    # Connect the dots: fit a linear model between the two points
    thisModel <- lm(Conc ~ Dis);
    as.numeric(predict(thisModel,newdata=data.frame(Dis=0)));
  }
  else {
    ## Fit a non-linear exponential decay
    concn.ss <- nls(Conc ~
negexp(Dis,B0,B1,theta),algorithm="plinear",control=list(maxiter =
500),trace=F);
    summary(concn.ss);
    predict(concn.ss,newdata=data.frame(Dis=0))[1];
  }
}
```

```

edgeConcPort <- function(Conc, Dis) {
  ## If either of the concentration valid numbers or distance valid numbers
  are less than 2, then connect the dots
  if(countNonZero(Conc) <= 2 | countNonZero(Dis) <= 2) {
    # Connect the dots: fit a linear model between the two points
    as.numeric(predict(thisModel,newdata=data.frame(Dis=0)));
  }
  else {
    ## Fit a non-linear exponential decay
    concn.ss <- nls(Conc ~
negexp(Dis,B0,B1,theta),algorithm="port",control=list(maxiter =
500),trace=F);
    predict(concn.ss,newdata=data.frame(Dis=0))[1];
  }
}

edgeConcLin <- function(Conc, Dis) {
  thisModel <- lm(Conc ~ Dis);
  as.numeric(predict(thisModel,newdata=data.frame(Dis=0)));
}

## Trellis graphics settings
graphics.off()
require(lattice)
require(RColorBrewer)
require(grid)
trellis.device()
cols <- c(brewer.pal(9,"Set1"),brewer.pal(8,"Dark2"))
sps <- trellis.par.get("superpose.symbol");
sps$col <- "black"
sps$pch <- 21:25
sps$fill <- cols
trellis.par.set("superpose.symbol",sps)

spl <- trellis.par.get("superpose.line")
spl$col <- brewer.pal(8,"Dark2");
spl$lty <- 1:6;
spl$lwd <- 2;
trellis.par.set("superpose.line",spl);
bringToTop(,stay=TRUE);

#*****
## Filename: Query edge distances.r
## Script purpose: Load RODBC and connect to the access database to extract
distance and concentration data.
#*****

library(RODBC);
connect = odbcConnectAccess("Concentration Studies.mdb");

# *****
# Collect observed concentrations from the concentration database.
# This set of queries collects the distances relative to the edge of road.
# *****

```



```

Alkanes <- sqlQuery(connect, "SELECT AlkaneMeasurements.Identifier,
AlkaneMeasurements.Bg, AlkaneMeasurements.Meas1, AlkaneMeasurements.Meas2,
AlkaneMeasurements.Meas3, AlkaneMeasurements.Meas4, AlkaneMeasurements.Meas5,
AlkaneMeasurements.Meas6, AlkaneMeasurements.Meas7, AlkaneMeasurements.Meas8,
AlkaneMeasurements.Meas9,[02Distances].NDist1, [02Distances].NDist2,
[02Distances].NDist3, [02Distances].NDist4, [02Distances].NDist5,
[02Distances].NDist6, [02Distances].NDist7, [02Distances].NDist8,
[02Distances].NDist9 FROM AlkaneMeasurements INNER JOIN [02Distances] ON
AlkaneMeasurements.Identifier=[02Distances].Identifier;");
BC <- sqlQuery(connect, "SELECT BCmeasurements.Identifier, BCmeasurements.Bg,
BCmeasurements.Meas1, BCmeasurements.Meas2, BCmeasurements.Meas3,
BCmeasurements.Meas4, BCmeasurements.Meas5, BCmeasurements.Meas6,
BCmeasurements.Meas7, BCmeasurements.Meas8,
BCmeasurements.Meas9,[02Distances].NDist1, [02Distances].NDist2,
[02Distances].NDist3, [02Distances].NDist4, [02Distances].NDist5,
[02Distances].NDist6, [02Distances].NDist7, [02Distances].NDist8,
[02Distances].NDist9 FROM BCmeasurements INNER JOIN [02Distances] ON
BCmeasurements.Identifier=[02Distances].Identifier;");
Benzene <- sqlQuery(connect, "SELECT BenzeneMeasurements.Identifier,
BenzeneMeasurements.Bg, BenzeneMeasurements.Meas1, BenzeneMeasurements.Meas2,
BenzeneMeasurements.Meas3, BenzeneMeasurements.Meas4,
BenzeneMeasurements.Meas5, BenzeneMeasurements.Meas6,
BenzeneMeasurements.Meas7, BenzeneMeasurements.Meas8,
BenzeneMeasurements.Meas9, [02Distances].NDist1, [02Distances].NDist2,
[02Distances].NDist3, [02Distances].NDist4, [02Distances].NDist5,
[02Distances].NDist6,
[02Distances].NDist7,[02Distances].NDist8,[02Distances].NDist9 FROM
BenzeneMeasurements INNER JOIN [02Distances] ON
BenzeneMeasurements.Identifier=[02Distances].Identifier;");
BS <- sqlQuery(connect, "SELECT BSmeasurements.Identifier, BSmeasurements.Bg,
BSmeasurements.Meas1, BSmeasurements.Meas2, BSmeasurements.Meas3,
BSmeasurements.Meas4, BSmeasurements.Meas5, BSmeasurements.Meas6,
BSmeasurements.Meas7, BSmeasurements.Meas8,
BSmeasurements.Meas9,[02Distances].NDist1, [02Distances].NDist2,
[02Distances].NDist3, [02Distances].NDist4, [02Distances].NDist5,
[02Distances].NDist6,
[02Distances].NDist7,[02Distances].NDist8,[02Distances].NDist9 FROM
BSmeasurements INNER JOIN [02Distances] ON
BSmeasurements.Identifier=[02Distances].Identifier;");
CO <- sqlQuery(connect, "SELECT COmeasurements.Identifier, COmeasurements.Bg,
COmeasurements.Meas1, COmeasurements.Meas2, COmeasurements.Meas3,
COmeasurements.Meas4, COmeasurements.Meas5, COmeasurements.Meas6,
COmeasurements.Meas7,COmeasurements.Meas8, COmeasurements.Meas9,
[02Distances].NDist1, [02Distances].NDist2, [02Distances].NDist3,
[02Distances].NDist4, [02Distances].NDist5, [02Distances].NDist6,
[02Distances].NDist7,[02Distances].NDist8,[02Distances].NDist9 FROM
COmeasurements INNER JOIN [02Distances] ON
COmeasurements.Identifier=[02Distances].Identifier;");
CO_2 <- sqlQuery(connect, "SELECT CO2measurements.Identifier,
CO2measurements.Bg, CO2measurements.Meas1, CO2measurements.Meas2,
CO2measurements.Meas3, CO2measurements.Meas4, CO2measurements.Meas5,
CO2measurements.Meas6, CO2measurements.Meas7, CO2measurements.Meas8,
CO2measurements.Meas9,[02Distances].NDist1, [02Distances].NDist2,
[02Distances].NDist3, [02Distances].NDist4, [02Distances].NDist5,
[02Distances].NDist6,
[02Distances].NDist7,[02Distances].NDist8,[02Distances].NDist9 FROM

```

```

CO2measurements INNER JOIN [02Distances] ON
CO2measurements.Identifier=[02Distances].Identifier;");
EC <- sqlQuery(connect, "SELECT ECmeasurements.Identifier, ECmeasurements.Bg,
ECmeasurements.Meas1, ECmeasurements.Meas2, ECmeasurements.Meas3,
ECmeasurements.Meas4, ECmeasurements.Meas5, ECmeasurements.Meas6,
ECmeasurements.Meas7, ECmeasurements.Meas8,
ECmeasurements.Meas9, [02Distances].NDist1, [02Distances].NDist2,
[02Distances].NDist3, [02Distances].NDist4, [02Distances].NDist5,
[02Distances].NDist6,
[02Distances].NDist7, [02Distances].NDist8, [02Distances].NDist9 FROM
ECmeasurements INNER JOIN [02Distances] ON
ECmeasurements.Identifier=[02Distances].Identifier;");
FineCoarse <- sqlQuery(connect, "SELECT FineCoarseMeasurements.Identifier,
FineCoarseMeasurements.Bg, FineCoarseMeasurements.Meas1,
FineCoarseMeasurements.Meas2, FineCoarseMeasurements.Meas3,
FineCoarseMeasurements.Meas4, FineCoarseMeasurements.Meas5,
FineCoarseMeasurements.Meas6, FineCoarseMeasurements.Meas7,
FineCoarseMeasurements.Meas8,
FineCoarseMeasurements.Meas9, [02Distances].NDist1, [02Distances].NDist2,
[02Distances].NDist3, [02Distances].NDist4, [02Distances].NDist5,
[02Distances].NDist6,
[02Distances].NDist7, [02Distances].NDist8, [02Distances].NDist9 FROM
FineCoarseMeasurements INNER JOIN [02Distances] ON
FineCoarseMeasurements.Identifier=[02Distances].Identifier;");
#HOC <- sqlQuery(connect, "SELECT HOCmeasurements.Identifier,
HOCmeasurements.Bg, HOCmeasurements.Meas1, HOCmeasurements.Meas2,
HOCmeasurements.Meas3, HOCmeasurements.Meas4, HOCmeasurements.Meas5,
HOCmeasurements.Meas6, HOCmeasurements.Meas7, HOCmeasurements.Meas8,
HOCmeasurements.Meas9, [02Distances].NDist1, [02Distances].NDist2,
[02Distances].NDist3, [02Distances].NDist4, [02Distances].NDist5,
[02Distances].NDist6,
[02Distances].NDist7, [02Distances].NDist8, [02Distances].NDist9 FROM
HOCmeasurements INNER JOIN [02Distances] ON
HOCmeasurements.Identifier=[02Distances].Identifier;");
#LOC <- sqlQuery(connect, "SELECT LOCmeasurements.Identifier,
LOCmeasurements.Bg, LOCmeasurements.Meas1, LOCmeasurements.Meas2,
LOCmeasurements.Meas3, LOCmeasurements.Meas4, LOCmeasurements.Meas5,
LOCmeasurements.Meas6, LOCmeasurements.Meas7, LOCmeasurements.Meas8,
LOCmeasurements.Meas9, [02Distances].NDist1, [02Distances].NDist2,
[02Distances].NDist3, [02Distances].NDist4, [02Distances].NDist5,
[02Distances].NDist6,
[02Distances].NDist7, [02Distances].NDist8, [02Distances].NDist9 FROM
LOCmeasurements INNER JOIN [02Distances] ON
LOCmeasurements.Identifier=[02Distances].Identifier;");
Metals <- sqlQuery(connect, "SELECT MetalsMeasurements.Identifier,
MetalsMeasurements.Bg, MetalsMeasurements.Meas1, MetalsMeasurements.Meas2,
MetalsMeasurements.Meas3, MetalsMeasurements.Meas4, MetalsMeasurements.Meas5,
MetalsMeasurements.Meas6, MetalsMeasurements.Meas7, MetalsMeasurements.Meas8,
MetalsMeasurements.Meas9, [02Distances].NDist1, [02Distances].NDist2,
[02Distances].NDist3, [02Distances].NDist4, [02Distances].NDist5,
[02Distances].NDist6,
[02Distances].NDist7, [02Distances].NDist8, [02Distances].NDist9 FROM
MetalsMeasurements INNER JOIN [02Distances] ON
MetalsMeasurements.Identifier=[02Distances].Identifier;");
NO <- sqlQuery(connect, "SELECT NOmeasurements.Identifier, NOmeasurements.Bg,
NOmeasurements.Meas1, NOmeasurements.Meas2, NOmeasurements.Meas3,
NOmeasurements.Meas4, NOmeasurements.Meas5, NOmeasurements.Meas6,

```

```

NOmeasurements.Meas7, NOmeasurements.Meas8,
NOmeasurements.Meas9,[02Distances].NDist1, [02Distances].NDist2,
[02Distances].NDist3, [02Distances].NDist4, [02Distances].NDist5,
[02Distances].NDist6,
[02Distances].NDist7,[02Distances].NDist8,[02Distances].NDist9 FROM
NOmeasurements INNER JOIN [02Distances] ON
NOmeasurements.Identifier=[02Distances].Identifier;");
nonAlkanes <- sqlQuery(connect, "SELECT NonAlkanesMeasurements.Identifier,
NonAlkanesMeasurements.Bg, NonAlkanesMeasurements.Meas1,
NonAlkanesMeasurements.Meas2, NonAlkanesMeasurements.Meas3,
NonAlkanesMeasurements.Meas4, NonAlkanesMeasurements.Meas5,
NonAlkanesMeasurements.Meas6, NonAlkanesMeasurements.Meas7,
NonAlkanesMeasurements.Meas8,
NonAlkanesMeasurements.Meas9,[02Distances].NDist1, [02Distances].NDist2,
[02Distances].NDist3, [02Distances].NDist4, [02Distances].NDist5,
[02Distances].NDist6, [02Distances].NDist7, [02Distances].NDist8,
[02Distances].NDist9 FROM NonAlkanesMeasurements INNER JOIN [02Distances] ON
NonAlkanesMeasurements.Identifier=[02Distances].Identifier;");
NO2 <- sqlQuery(connect, "SELECT NO2measurements.Identifier,
NO2measurements.Bg, NO2measurements.Meas1, NO2measurements.Meas2,
NO2measurements.Meas3, NO2measurements.Meas4, NO2measurements.Meas5,
NO2measurements.Meas6, NO2measurements.Meas7, NO2measurements.Meas8,
NO2measurements.Meas9,[02Distances].NDist1, [02Distances].NDist2,
[02Distances].NDist3, [02Distances].NDist4, [02Distances].NDist5,
[02Distances].NDist6,
[02Distances].NDist7,[02Distances].NDist8,[02Distances].NDist9 FROM
NO2measurements INNER JOIN [02Distances] ON
NO2measurements.Identifier=[02Distances].Identifier;");
NOx <- sqlQuery(connect, "SELECT NOxMeasurements.Identifier,
NOxMeasurements.Bg, NOxMeasurements.Meas1, NOxMeasurements.Meas2,
NOxMeasurements.Meas3, NOxMeasurements.Meas4, NOxMeasurements.Meas5,
NOxMeasurements.Meas6, NOxMeasurements.Meas7, NOxMeasurements.Meas8,
NOxMeasurements.Meas9,[02Distances].NDist1, [02Distances].NDist2,
[02Distances].NDist3, [02Distances].NDist4, [02Distances].NDist5,
[02Distances].NDist6,
[02Distances].NDist7,[02Distances].NDist8,[02Distances].NDist9 FROM
NOxMeasurements INNER JOIN [02Distances] ON
NOxMeasurements.Identifier=[02Distances].Identifier;");
OC <- sqlQuery(connect, "SELECT OCmeasurements.Identifier, OCmeasurements.Bg,
OCmeasurements.Meas1, OCmeasurements.Meas2, OCmeasurements.Meas3,
OCmeasurements.Meas4, OCmeasurements.Meas5, OCmeasurements.Meas6,
OCmeasurements.Meas7, OCmeasurements.Meas8,
OCmeasurements.Meas9,[02Distances].NDist1, [02Distances].NDist2,
[02Distances].NDist3, [02Distances].NDist4, [02Distances].NDist5,
[02Distances].NDist6,
[02Distances].NDist7,[02Distances].NDist8,[02Distances].NDist9 FROM
OCmeasurements INNER JOIN [02Distances] ON
OCmeasurements.Identifier=[02Distances].Identifier;");
Ozone <- sqlQuery(connect, "SELECT OzoneMeasurements.Identifier,
OzoneMeasurements.Bg, OzoneMeasurements.Meas1, OzoneMeasurements.Meas2,
OzoneMeasurements.Meas3, OzoneMeasurements.Meas4, OzoneMeasurements.Meas5,
OzoneMeasurements.Meas6, OzoneMeasurements.Meas7, OzoneMeasurements.Meas8,
OzoneMeasurements.Meas9,[02Distances].NDist1, [02Distances].NDist2,
[02Distances].NDist3, [02Distances].NDist4, [02Distances].NDist5,
[02Distances].NDist6,
[02Distances].NDist7,[02Distances].NDist8,[02Distances].NDist9 FROM

```

```

OzoneMeasurements INNER JOIN [02Distances] ON
OzoneMeasurements.Identifier=[02Distances].Identifier;");
PM10abs <- sqlQuery(connect, "SELECT PM10AbsCoeffMeasurements.Identifier,
PM10AbsCoeffMeasurements.Bg, PM10AbsCoeffMeasurements.Meas1,
PM10AbsCoeffMeasurements.Meas2, PM10AbsCoeffMeasurements.Meas3,
PM10AbsCoeffMeasurements.Meas4, PM10AbsCoeffMeasurements.Meas5,
PM10AbsCoeffMeasurements.Meas6, PM10AbsCoeffMeasurements.Meas7,
PM10AbsCoeffMeasurements.Meas8,
PM10AbsCoeffMeasurements.Meas9,[02Distances].NDist1, [02Distances].NDist2,
[02Distances].NDist3, [02Distances].NDist4, [02Distances].NDist5,
[02Distances].NDist6,
[02Distances].NDist7,[02Distances].NDist8,[02Distances].NDist9 FROM
PM10AbsCoeffMeasurements INNER JOIN [02Distances] ON
PM10AbsCoeffMeasurements.Identifier=[02Distances].Identifier;");
PM10 <- sqlQuery(connect, "SELECT PM10Measurements.Identifier,
PM10Measurements.Bg, PM10Measurements.Meas1, PM10Measurements.Meas2,
PM10Measurements.Meas3, PM10Measurements.Meas4, PM10Measurements.Meas5,
PM10Measurements.Meas6, PM10Measurements.Meas7, PM10Measurements.Meas8,
PM10Measurements.Meas9,[02Distances].NDist1, [02Distances].NDist2,
[02Distances].NDist3, [02Distances].NDist4, [02Distances].NDist5,
[02Distances].NDist6,
[02Distances].NDist7,[02Distances].NDist8,[02Distances].NDist9 FROM
PM10Measurements INNER JOIN [02Distances] ON
PM10Measurements.Identifier=[02Distances].Identifier;");
PM2_5abs <- sqlQuery(connect, "SELECT PM25AbsCoeffMeasurements.Identifier,
PM25AbsCoeffMeasurements.Bg, PM25AbsCoeffMeasurements.Meas1,
PM25AbsCoeffMeasurements.Meas2, PM25AbsCoeffMeasurements.Meas3,
PM25AbsCoeffMeasurements.Meas4, PM25AbsCoeffMeasurements.Meas5,
PM25AbsCoeffMeasurements.Meas6, PM25AbsCoeffMeasurements.Meas7,
PM25AbsCoeffMeasurements.Meas8,
PM25AbsCoeffMeasurements.Meas9,[02Distances].NDist1, [02Distances].NDist2,
[02Distances].NDist3, [02Distances].NDist4, [02Distances].NDist5,
[02Distances].NDist6,
[02Distances].NDist7,[02Distances].NDist8,[02Distances].NDist9 FROM
PM25AbsCoeffMeasurements INNER JOIN [02Distances] ON
PM25AbsCoeffMeasurements.Identifier=[02Distances].Identifier;");
PM2_5 <- sqlQuery(connect, "SELECT PM25Measurements.Identifier,
PM25Measurements.Bg, PM25Measurements.Meas1, PM25Measurements.Meas2,
PM25Measurements.Meas3, PM25Measurements.Meas4, PM25Measurements.Meas5,
PM25Measurements.Meas6, PM25Measurements.Meas7, PM25Measurements.Meas8,
PM25Measurements.Meas9,[02Distances].NDist1, [02Distances].NDist2,
[02Distances].NDist3, [02Distances].NDist4, [02Distances].NDist5,
[02Distances].NDist6,
[02Distances].NDist7,[02Distances].NDist8,[02Distances].NDist9 FROM
PM25Measurements INNER JOIN [02Distances] ON
PM25Measurements.Identifier=[02Distances].Identifier;");
Sulfur <- sqlQuery(connect, "SELECT SulfurMeasurements.Identifier,
SulfurMeasurements.Bg, SulfurMeasurements.Meas1, SulfurMeasurements.Meas2,
SulfurMeasurements.Meas3, SulfurMeasurements.Meas4, SulfurMeasurements.Meas5,
SulfurMeasurements.Meas6, SulfurMeasurements.Meas7, SulfurMeasurements.Meas8,
SulfurMeasurements.Meas9,[02Distances].NDist1, [02Distances].NDist2,
[02Distances].NDist3, [02Distances].NDist4, [02Distances].NDist5,
[02Distances].NDist6,
[02Distances].NDist7,[02Distances].NDist8,[02Distances].NDist9 FROM
SulfurMeasurements INNER JOIN [02Distances] ON
SulfurMeasurements.Identifier=[02Distances].Identifier;");

```

```

TSP <- sqlQuery(connect, "SELECT TSPMeasurements.Identifier,
TSPMeasurements.Bg, TSPMeasurements.Meas1, TSPMeasurements.Meas2,
TSPMeasurements.Meas3, TSPMeasurements.Meas4, TSPMeasurements.Meas5,
TSPMeasurements.Meas6, TSPMeasurements.Meas7, TSPMeasurements.Meas8,
TSPMeasurements.Meas9,[02Distances].NDist1, [02Distances].NDist2,
[02Distances].NDist3, [02Distances].NDist4, [02Distances].NDist5,
[02Distances].NDist6,
[02Distances].NDist7,[02Distances].NDist8,[02Distances].NDist9 FROM
TSPMeasurements INNER JOIN [02Distances] ON
TSPMeasurements.Identifier=[02Distances].Identifier;");
Ultrafine <- sqlQuery(connect, "SELECT UltrafineMeasurements.Identifier,
UltrafineMeasurements.Bg, UltrafineMeasurements.Meas1,
UltrafineMeasurements.Meas2, UltrafineMeasurements.Meas3,
UltrafineMeasurements.Meas4, UltrafineMeasurements.Meas5,
UltrafineMeasurements.Meas6, UltrafineMeasurements.Meas7,
UltrafineMeasurements.Meas8,
UltrafineMeasurements.Meas9,[02Distances].NDist1, [02Distances].NDist2,
[02Distances].NDist3, [02Distances].NDist4, [02Distances].NDist5,
[02Distances].NDist6,
[02Distances].NDist7,[02Distances].NDist8,[02Distances].NDist9 FROM
UltrafineMeasurements INNER JOIN [02Distances] ON
UltrafineMeasurements.Identifier=[02Distances].Identifier;");
UltrafineFine <- sqlQuery(connect, "SELECT
UltrafineFineMeasurements.Identifier, UltrafineFineMeasurements.Bg,
UltrafineFineMeasurements.Meas1, UltrafineFineMeasurements.Meas2,
UltrafineFineMeasurements.Meas3, UltrafineFineMeasurements.Meas4,
UltrafineFineMeasurements.Meas5, UltrafineFineMeasurements.Meas6,
UltrafineFineMeasurements.Meas7, UltrafineFineMeasurements.Meas8,
UltrafineFineMeasurements.Meas9,[02Distances].NDist1, [02Distances].NDist2,
[02Distances].NDist3, [02Distances].NDist4, [02Distances].NDist5,
[02Distances].NDist6,
[02Distances].NDist7,[02Distances].NDist8,[02Distances].NDist9 FROM
UltrafineFineMeasurements INNER JOIN [02Distances] ON
UltrafineFineMeasurements.Identifier=[02Distances].Identifier;");

# Close database connection
odbcClose(connect);

#####
## Filename: Data Prep 1.r
## Script purpose: Take the raw data from the database and transform it into
a transposed dataframe for each pollutant, containing the edge interpolated
values using both exponential and linear fits.
#####

## Definitions used for charts and data frame labels.

## Combine all measures of elemental carbon into one
## Use this definition of BC for all normalized plots
BC <- rbind(BC,EC,PM2_5abs,PM10abs,BS)

## Use this definition of BC for all unnormalized plots
#BC <- rbind(BC,EC)
TSP[,2:11] <- TSP[,2:11]*0.55
PM10 <- rbind(PM10,TSP)

```

```

pollutants <-
c("Alkanes", "BC", "Benzene", "CO", "CO_2", "FineCoarse", "Metals", "NO", "nonAlkanes",
  "NO2", "NOx", "OC",
  "Ozone", "PM10", "PM2_5", "Sulfur", "Ultrafine", "UltrafineFine");

classes <-
c("Gas", "Other", "Gas", "Gas", "Gas", "Number", "Other", "Gas", "Gas", "Gas", "Gas", "O",
  "ther", "Gas", "Mass", "Mass", "Other", "Number", "Number");

measurementUnits <-
c(" (ppb) ", " (ug/m3) ", " (ug/m3) ", " (ppm) ", " (ppm) ", " (/cm3) ", " (ug/m2/d) ", " (ppb) ", " (",
  "ppb) ", " (ppb) ", " (ppb) ", " (ug/m3) ", " (ppb) ", " (ug/m3) ", " (ug/m3) ", " (ug/m3) ", " (/cm3)",
  " (/cm3) ");

newPollutants <- c("Alkanes", "Elemental carbon", "Benzene", "CO", "CO2", "Fine
  particle no.", "Metal deposition", "NO", "Non-Alkane
  hydrocarbons", "NO2", "NOx", "Organic carbon",
  "Ozone", "PM10", "PM2.5", "Sulfur", "UF1 particle no.", "UF2 particle no.");

newNames <-
c("Bg", "Meas0", "Meas1", "Meas2", "Meas3", "Meas4", "Meas5", "Meas6", "Meas7", "Meas8",
  "Meas9", "NDist0", "NDist1", "NDist2", "NDist3", "NDist4", "NDist5", "NDist6", "NDi",
  "st7", "NDist8", "NDist9", "hasBG", "hasTrend");

## *****

## *****
## This subroutine generates one dataframe for each pollutant entitled
## "Pollutant".t It contains all distance/concentration data, and a boolean
## variable indicating the presence or absence of background data and/or a
## trend.

## AADT <-
data.frame("Identifier"=traffic[, "Identifier"], "AADT"=traffic[, "AADT"]);
for(i in 1:length(pollutants)) { #c(17,19,22,23)) {
  #if(length(na.omit(match(i,c(3,5,7,9,11,13,15,17,19,21,23))))==1) {
pdf(file="Rplots.pdf"); par(mfrow=c(1,2));}

## Attach traffic data
## eval(parse(text=paste(pollutants[i], " <-
merge(" , pollutants[i], " , AADT) ; " , sep=""))));

  ID <- as.vector(eval(parse(text=pollutants[i]))$Identifier);

## Transposing the data creates vectors of observed concentrations and
## distances from the road. Strip study names since this would change the mode
## of the final data frame.
  assign(paste(pollutants[i], ".t", sep=""),
as.data.frame(t(data.frame(eval(parse(text=pollutants[i]))[, -1]))));

## Which series have background values? bind to the data frame for later use
  eval(parse(text=paste(pollutants[i], ".t <-
rbind(" , pollutants[i], ".t , " , pollutants[i], ".t[\"Bg\", ]>0) ; " , sep=""))));

## Which series indicate a trend? bind to the data frame for later use
  eval(parse(text=paste("temp <- rep(0, ncol(" , pollutants[i], ".t))" , sep=""))));

```

```

    eval(parse(text=paste("for(x in 1:ncol(", pollutants[i], ".t)) { temp[x] <-
countNonZero(", pollutants[i], ".t[2:10,x])>1; }", sep=""))));
    eval(parse(text=paste(pollutants[i], ".t <- rbind(", pollutants[i], ".t,
temp);", sep=""))));

## curr is a temporary placeholder for the current data frame. It keeps us
from using eval, parse constructions.
    assign("curr",
as.data.frame(eval(parse(text=paste(pollutants[i], ".t", sep="")))));

##Add rows representing a distance point for the edge of road, and the
concentration at that point (they will be used later).
    eval(parse(text=(paste(pollutants[i], ".t <-
as.data.frame(", pollutants[i], ".t[c(1,1:10,10:(nrow(", pollutants[i], ".t))),]
", sep=""))));

    eval(parse(text=paste("names(", pollutants[i], ".t", ") <- ID", sep=""))));
    eval(parse(text=paste(pollutants[i], ".t <-
data.frame(", pollutants[i], ".t, check.names=TRUE)", sep=""))));
    eval(parse(text=(paste(pollutants[i], ".t[2,] <-
rep(0, ncol(", pollutants[i], ".t)", sep=""))));
    eval(parse(text=(paste(pollutants[i], ".t[12,] <-
rep(0, ncol(", pollutants[i], ".t)", sep=""))));
    eval(parse(text=(paste("row.names(", pollutants[i], ".t) <-
newNames", sep=""))));
    rm(curr);
}

## Interpolate the edge of road concentration for those studies for which a
trend exists
pollutantsExp <-
c("BC", "Benzene", "CO", "Metals", "NO", "NO2", "NOx", "Ozone", "Sulfur", "Ultrafine",
"UltrafineFine");
for(i in 1:length(pollutantsExp)) {
  assign("curr", eval(parse(text=paste(pollutantsExp[i], ".t", sep=""))));
  for(j in 1:ncol(curr)) {
    print(i); print(j);
    Concentrations <- curr[3:11,j];
    Distances <- curr[13:21,j];
    ## Not possible to fit these ones by plinear algorithm
    if((i == 1 & (j == 37 | j == 40 | j == 71 | j == 74)) | (i == 2 & (j == 3
| j == 4)))
      try(eval(parse(text=paste(pollutantsExp[i], ".t[\"Meas0\",j] <-
", edgeConcPort(Concentrations, Distances), sep=""))))
    else {
      if(i == 2 & j == 2)
        try(eval(parse(text=paste(pollutantsExp[i], ".t[\"Meas0\",j] <-
", edgeConcLin(Concentrations, Distances), sep=""))))
      else {
        if(curr["hasTrend",j] == 1)
          if(curr["NDist1",j] <= 1)
            eval(parse(text=paste(pollutantsExp[i], ".t[\"Meas0\",j] <-
", pollutantsExp[i], ".t[\"Meas1\",j]", sep=""))))
          else
            try(eval(parse(text=paste(pollutantsExp[i], ".t[\"Meas0\",j] <-
", edgeConcExp(Concentrations, Distances), sep=""))))
      }
    }
  }
}

```

```

    }

#Fit non-alkanes using the "port" nonlinear algorithm
curr <- nonAlkanes.t
for(j in 1:ncol(curr)) {
  print(i); print(j);
  Concentrations <- curr[3:11,j];
  Distances <- curr[13:21,j];
  if(j == 6 | j == 12) {
    thisModel <- lm(Concentrations[1:2] ~ Distances[1:2]);
    nonAlkanes.t["Meas0",j] <- as.numeric(thisModel$coefficients[1])
  }
  else {
    if(curr["hasTrend",j] == 1)
      if(curr["NDist1",j] <= 1)
        nonAlkanes.t["Meas0",j] <- nonAlkanes.t["Meas1",j]
    else
      try(nonAlkanes.t["Meas0",j] <-
edgeConcPort(Concentrations,Distances))
  }
}

pollutantsLin <- c("Alkanes","CO_2","FineCoarse","OC","PM10","PM2_5");
for(i in 1:length(pollutantsLin)) {
  assign("curr", eval(parse(text=paste(pollutantsLin[i],".t",sep=""))));
  for(j in 1:ncol(curr)) {
    print(i); print(j);
    Concentrations <- curr[3:11,j];
    Distances <- curr[13:21,j];
    if(curr["hasTrend",j] == 1)
      if(curr["NDist1",j] <= 1)
        eval(parse(text=paste(pollutantsLin[i],".t\\\"Meas0\\\",j] <-
",pollutantsLin[i],".t\\\"Meas1\\\",j]",sep="")))
      else
        try(eval(parse(text=paste(pollutantsLin[i],".t\\\"Meas0\\\",j] <-
",edgeConcLin(Concentrations,Distances),sep=""))));
  }
}

#*****
## Filename: Data Prep 2.r
## Script purpose: Transform the transposed data frames into long form
datasets normalized to edge, and normalized to background
#*****

# *****
# TRANSFORMATION TO LONG FORM
# *****

## Create a large matrix containing all the data, to facilitate modeling and
plotting.
# *****
dataSet <- data.frame();
for(i in 1:length(pollutants)) {
  assign("curr",
as.data.frame(eval(parse(text=paste(pollutants[i],".t",sep=""))));
  curr[is.na(curr)] <- 0;

```



```

for(j in 1:dim(curr)[2]) {
  concentrations <- curr[3:11,j] > 0;
  distances <- curr[13:21,j] > 0;
  for(k in 3:11) {
    if(concentrations[k-2] == distances[k-2] & curr[k,j] > 0) {
      dataSet <- rbind(dataSet,
data.frame(Pollutant=newPollutants[i],ID=names(curr)[j],Concentration=curr[k,
j],Distance=curr[k+10,j], hasBG=curr["hasBG",j],
hasTrend=curr["hasTrend",j]));
    }
  }
  rm(curr);
}
# *****

# Create a parallel dataSet with only trend measurements, normalized to edge
of road concentrations
# *****
normToEdge_dataSet <- data.frame();
for(i in 1:length(pollutants)) {
  assign("curr",
as.data.frame(eval(parse(text=paste(pollutants[i],".t",sep="")))));
  curr[is.na(curr)] <- 0;
  for(j in 1:dim(curr)[2]) {
    concentrations <- curr[2:11,j] > 0;
    distances <- curr[12:21,j] >= 0;
    for(k in 2:11) {
      if(concentrations[k-1] == distances[k-1] & curr[k,j] > 0 &
curr["hasTrend",j] == 1) {
        normToEdge_dataSet <- rbind(normToEdge_dataSet,
data.frame(Pollutant=newPollutants[i],Class=classes[i],ID=names(curr)[j],Conc
entration=curr[k,j],Distance=curr[k+10,j],AADT=curr["AADT",j]));
      }
    }
  }
  rm(curr);
}
zeroIndex <- normToEdge_dataSet$Distance==0.0;
normToEdge_dataSet <-
cbind(normToEdge_dataSet,data.frame("Normalized"=rep(0,nrow(normToEdge_dataSe
t))));
for(i in 1:length(zeroIndex)) {
  if(zeroIndex[i]) {
    trueIndex <- i;
    normToEdge_dataSet[i,"Normalized"] <- 1.0;
  }
  if(!zeroIndex[i])
    normToEdge_dataSet[i,"Normalized"] <-
normToEdge_dataSet[i,"Concentration"]/normToEdge_dataSet[trueIndex,"Concentra
tion"];
}

# Create a parallel dataSet with only trend measurements, normalized to
BACKGROUND
# *****

```

```

BG_ds <- data.frame();
for(i in 1:length(pollutants)) {
  assign("curr",
as.data.frame(eval(parse(text=paste(pollutants[i],".t",sep="")))));
  curr[is.na(curr)] <- 0;
  for(j in 1:dim(curr)[2]) {
    ## Indexing vectors to find location of data
    concentrations <- curr[3:11,j] > 0;
    distances <- curr[12:21,j] >= 0;
    for(k in 3:11) {
      if(concentrations[k-2] == distances[k-2] & curr[k,j] > 0 &
curr["hasBG",j] == 1) {
        BG_ds <- rbind(BG_ds,
data.frame(Pollutant=newPollutants[i],Class=classes[i],ID=names(curr)[j],Back
ground=curr[1,j],Concentration=curr[k,j],Distance=curr[k+10,j],AADT=curr["AAD
T",j]));
      }
    }
  }
  rm(curr);
}

## Generate a column with data normalized to the background concentration
BG_ds <- cbind(BG_ds,data.frame("Normalized"=rep(0,nrow(BG_ds))));
BG_ds[, "Normalized"] <- BG_ds[, "Concentration"]/BG_ds[, "Background"]

#####
## Filename: Trellis charts_background.r
## Script purpose: Create plots from the background normalized data
#####

allPollutantsbyClass <- with(BG_ds,
  xyplot(
    Normalized ~ Distance, as.table=TRUE,
    aspect = 1.3,
    subscripts=TRUE,
    layout = c(2,1),
    group.list = list(Pollutant,Class,Pollutant,Class),
    xlab=list("Distance from edge
(m)",fontfamily="Calibri",fontface=2,cex=1.0),
    ylab=list("Pollutant concentration normalized to
background",fontfamily="Calibri",fontface=2,cex=1.0),
    scales = list(fontfamily="Calibri"),
    prepanel = function(x, y, ...) {
      list(xlim=range(0,1000), ylim = range(0, 10), dx = NULL, dy = NULL)
    },
    panel = function(...,x,y,group.list) {
      panel.xyplot(...,x,y,groups = group.list[[panel.number()]],);
      panel.abline(1,0,col="gray");
      panel.abline(1.5,0,col="black");
    },
    legend = list(
      inside = list(corner=c(1,1),x=0.5, y=0.96,
        fun="draw.key",
        args = list(
          key = list(

```

```

        transparent=T, fontfamily="Calibri", cex =0.8, border=0,
between=0.4, columns=1,
        points=Rows(sps, 1:length(unique(BG_ds$Pollutant))),
        text=list(as.character(unique(BG_ds$Pollutant))), draw =
FALSE)),
    inside = list(corner=c(1,1),x=0.92, y=0.8,
        fun="draw.key",
        args = list(
            key = list(
                transparent=T,fontfamily="Calibri", cex =0.8, border=0,
between=0.4, columns=1,
                points=Rows(sps, 1:length(unique(BG_ds$Class))),
                text=list(as.character(unique(BG_ds$Class))), draw = FALSE)))
        #Previous key, useful if parameters change:
as.vector(unique(normToEdge_dataSet$Pollutant)))
    )
);
allPollutantsbyClass[c(1,1)]

normalizedPanels <-xyplot(
    Normalized ~ Distance | Pollutant,
    data=BG_ds,
    as.table=TRUE,
    groups=ID,
    xlab=list("Distance from edge
(m)",fontfamily="Calibri",fontface=2,cex=1.0),
    ylab=list("Pollutant concentration normalized to
background",fontfamily="Calibri",fontface=2,cex=1.0),
    scales=list(relation="free",fontfamily="Calibri"),
    prepanel = function(x, y, ...) {
        list(xlim=range(0,400), ylim = range(0, y), dx = NULL, dy = NULL)
    },
    panel = function(x,y,...) {
        panel.xyplot(x,y,...);
        #panel.loess(x,y,...)
    }

#strip.custom(5,par.strip.text=list(fontfamily="Calibri",fontface=2,cex=0.65)
)

);

## Varies with distance
col1 <- trellis.par.get("strip.background")$col[1];
## Does not vary with distance
col2 <- trellis.par.get("strip.background")$col[2];
## Assign
cols <-
c(col2,col1,col1,col1,col2,col1,col1,col1,col2,col1,col2,col1,col2,col2,
col1,col1);

update(normalizedPanels[c(3,4,2,6,7,8,10,12,16,17,1,5,9,11,13,14,15)],
    strip = function(...,which.panel,bg,par.strip.text) {
        bg.col = cols;
        strip.default(..., which.panel = which.panel,
            bg = rep(bg.col, length = which.panel)[which.panel],

par.strip.text=list(fontfamily="Calibri",fontface=2,cex=0.65))

```

```

    },
    layout=c(5,4,1),
    skip=c(rep(F,15),T,F,T,F,T)
  )
# *****

index <- BG_ds$Pollutant=="Organic carbon" | (BG_ds$ID=="YZ2006" &
BG_ds$Pollutant=="UF1 particle no.") | (BG_ds$ID=="CR1981" &
BG_ds$Pollutant=="NO") | (BG_ds$ID=="CR1981" & BG_ds$Pollutant=="NOx") |
(BG_ds$ID=="CR1981" & BG_ds$Pollutant=="NO2") | BG_ds$Pollutant=="Sulfur"

newD <- BG_ds[!index,]
pointsD <- BG_ds[index,]
newD$Pollutant <- newD$Pollutant[, drop = TRUE]
pointsD$Pollutant <- pointsD$Pollutant[, drop = TRUE]
levels(pointsD$Pollutant) <- c(levels(pointsD$Pollutant), "Rodes and Holland
(1981) - NO", "Rodes and Holland (1981) - NOx", "Rodes and Holland (1981) -
NO2", "Zhu et al. (2006) - UF1")
pointsD$Pollutant[pointsD$Pollutant=="UF1 particle no."] = "Zhu et al. (2006)
- UF1"
pointsD$Pollutant[pointsD$Pollutant=="NO"] = "Rodes and Holland (1981) - NO"
pointsD$Pollutant[pointsD$Pollutant=="NOx"] = "Rodes and Holland (1981) -
NOx"
pointsD$Pollutant[pointsD$Pollutant=="NO2"] = "Rodes and Holland (1981) -
NO2"
pointsD$Pollutant <- pointsD$Pollutant[, drop = TRUE]

## Count each "n"--the number of observations for each pollutant to include
on the plot
numPols <- match(unique(newD$Pollutant), newD$Pollutant)
subIn <- 2:length(numPols)
numPols <- numPols[subIn] - numPols[subIn-1]
keyTxt <- c(paste(as.character(unique(newD$Pollutant)), "
(", numPols, ") ", sep=""))

## Append baseline line to legend lines
polLns <- Rows(spl, 1:length(unique(newD$Pollutant)))
polLns$alpha <- c(polLns$alpha, 1, 1)
polLns$col <- c(polLns$col, "black", "gray")
polLns$lty <- c(polLns$lty, 1, 1)
polLns$lwd <- c(polLns$lwd, 1, 2)

lfits_BG <- xyplot(Normalized ~ Distance, data = newD, groups=Pollutant,
  aspect=0.6,
  type="smooth", lwd=2.0, span=0.75, degree=1,
  xlab=list("Distance from edge
(m)", fontfamily="Calibri", fontface=2, cex=1.0),
  ylab=list("Pollutant concentration normalized to
background", fontfamily="Calibri", fontface=2, cex=1.0),
  scales=list(fontfamily="Calibri"),
  prepanel = function(x, y, ...) {
    list(xlim=range(0, 450), ylim = range(0, 5), dx = NULL, dy = NULL)
  },
  panel = function(...) {
    panel.superpose(...);
  },
  panel.groups = function(x, y, ...) {

```

```

panel.xyplot(x,y,...);
panel.abline(1,0,col="black");

panel.xyplot(pointsD$Distance,pointsD$Normalized,groups=pointsD$Pollutant,sub
scripts=seq_len(nrow(pointsD)))
  thisModel <- locfit(y ~ lp(x,nn=0.75,deg=1));
  preds <- predict(thisModel,newdata=data.frame(x=max(x):450));
  len <- which(preds >= 1);
  dists <- seq(from=max(x), length.out=length(len));
  if(max(x) < 450) llines(dists,preds[len],col="gray",lty=1,lwd=2)
},
legend = list(
  inside = list(corner=c(1,1),x=0.98, y=0.98,
    fun="draw.key",
    args = list(
      key = list(
        transparent=F,background="white",fontfamily="Calibri", cex =0.8,
border=0, columns=2,between=0.3, between.columns=0.3,
        lines=polLns,
        text=list(c(keyTxt,"Background concentration","Prediction lines"))
      )
    ),
  inside = list(corner=c(1,1),x=0.98, y=0.68,
    fun="draw.key",
    args = list(
      key = list(
        transparent=F,background="white",fontfamily="Calibri", cex =0.8,
border=0, columns=2,between=0.5,
        points=Rows(sps, 1:length(unique(pointsD$Pollutant))),
        text=list(levels(pointsD$Pollutant))
      )))
)

## Plot single panel version
lfits_BG;

## Plot two panel version with all gray lines in the bottom panel
update(lfits_BG[c(1,1)],as.table=TRUE,aspect=0.3,layout=c(1,2))

xyplot(Normalized ~ Distance, data = newD, groups=Pollutant,aspect=0.7,
  panel = panel.superpose,
  panel.groups = function(x,y,...) {
    if panel.number()==1{
      thisModel <- locfit(y ~ lp(x,nn=0.75,deg=1),maxk=500);

      llines(0:450,predict(thisModel,newdata=data.frame(x=0:450)),col="gray",lty=1,
      lwd=2)
    }
  })

#*****
## Filename: Trellis charts_edge.r
## Script purpose: Create plots from the edge normalized data
#*****

```

```

allPollutantsbyClass <- with(normToEdge_dataSet,
  xyplot(
    Normalized ~ Distance, as.table=TRUE,
    aspect = 1.5,
    subscripts=TRUE,
    layout = c(2,1),
    group.list = list(Pollutant,Class,Pollutant,Class),
    xlab=list("Distance from edge
(m)",fontfamily="Calibri",fontface=2,cex=1.0),
    ylab=list("Pollutant concentration normalized to
edge",fontfamily="Calibri",fontface=2,cex=1.0),
    scales = list(fontfamily="Calibri"),
    prepanel = function(x, y, ...) {
      list(xlim=range(0,450), ylim = range(0, 3), dx = NULL, dy = NULL)
    },
    panel = function(...,x,y,group.list) {
      panel.xyplot(...,x,y,groups = group.list[[panel.number()]],)
    },
    legend = list(
      inside = list(corner=c(1,1),x=0.48, y=0.96,
        fun="draw.key",
        args = list(
          key = list(
            transparent=T, fontfamily="Calibri", cex =0.7, border=0,
columns=2, between.columns=0.3,between=0.3,
            points=Rows(sps, 1:length(unique(normToEdge_dataSet$Pollutant))),
            text=list(as.character(unique(normToEdge_dataSet$Pollutant)))),
draw = FALSE)),
      inside = list(corner=c(1,1),x=0.85, y=0.9,
        fun="draw.key",
        args = list(
          key = list(
            transparent=T,fontfamily="Calibri", cex =0.7, border=0,
columns=2, between=0.3,
            points=Rows(sps, 1:length(unique(normToEdge_dataSet$Class))),
            text=list(as.character(unique(normToEdge_dataSet$Class))), draw
= FALSE)))
      #Previous key, useful if parameters change:
as.vector(unique(normToEdge_dataSet$Pollutant)))
    )
  );
allPollutantsbyClass[c(1,1)]

normalizedPanels <-xyplot(
  Normalized ~ Distance | Pollutant,
  data=normToEdge_dataSet,
  as.table=TRUE,
  groups=ID,
  xlab=list("Distance from edge
(m)",fontfamily="Calibri",fontface=2,cex=1.0),
  ylab=list("Pollutant concentration normalized to
edge",fontfamily="Calibri",fontface=2,cex=1.0),
  scales=list(fontfamily="Calibri"),
  prepanel = function(x, y, ...) {
    list(xlim=range(0,400), ylim = range(0, 2), dx = NULL, dy = NULL)
  },
  panel = function(x,y,...) {

```

```

        panel.xyplot(x,y,...);
        #panel.loess(x,y,...)
    }

#strip.custom(5,par.strip.text=list(fontfamily="Calibri",fontface=2,cex=0.65)
)

    );

    # Varies with distance
    col1 <- trellis.par.get("strip.background")$col[1];
    # Does not vary with distance
    col2 <- trellis.par.get("strip.background")$col[2];
    cols <-
c(col2,col1,col1,col1,col2,col1,col1,col1,col1,col1,col2,col1,col2,col2,col2,
col1,col1);

    update(normalizedPanels[c(2,3,4,6,7,8,9,10,12,16,17,1,5,11,13,14,15)],
        strip = function(...,which.panel,bg,par.strip.text) {
            bg.col = cols;
            strip.default(..., which.panel = which.panel,
                bg = rep(bg.col, length = which.panel)[which.panel],

par.strip.text=list(fontfamily="Calibri",fontface=2,cex=0.65))
        },
        layout=c(4,5,1),
        skip=c(rep(F,11),T,rep(F,7)))
# *****

## Eliminate the following pollutants--they cause the smoother to perform
erratically.
index <- normToEdge_dataSet$Pollutant=="Organic carbon" |
normToEdge_dataSet$Pollutant=="Sulfur" | normToEdge_dataSet$Pollutant=="Metal
deposition"

### *****

newD <- normToEdge_dataSet[!index,]
pointsD <- normToEdge_dataSet[index,]
newD$Pollutant <- newD$Pollutant[, drop = TRUE]
pointsD$Pollutant <- pointsD$Pollutant[, drop = TRUE]

## Count each "n"--the number of observations for each pollutant to include
on the plot
numPols <- match(unique(newD$Pollutant),newD$Pollutant)
subIn <- 2:length(numPols)
numPols <- numPols[subIn] - numPols[subIn-1]
keyTxt <- paste(levels(newD$Pollutant)," (" ,numPols,") ",sep="")

lfits <- xyplot(Normalized ~ Distance,data = newD, groups=Pollutant,
    aspect=0.7,
    type=c("smooth","g"),lwd=2.0,span=0.7,degree=1,
    xlab=list("Distance from edge
(m)",fontfamily="Calibri",fontface=2,cex=1.0),
    ylab=list("Pollutant concentration normalized to
edge",fontfamily="Calibri",fontface=2,cex=1.0),
    scales=list(fontfamily="Calibri"),
    prepanel = function(x, y, ...) {

```

```

        list(xlim=range(0,400), ylim = range(0, 1.5), dx = NULL, dy = NULL)
    },
    panel = function(...) {
        panel.superpose(...);
    },
    panel.groups = function(x,y,...) {
        panel.xyplot(x,y,...);

panel.xyplot(pointsD$Distance,pointsD$Normalized,groups=pointsD$Pollutant,sub
scripts=seq_len(nrow(pointsD)));
        panel.abline(0.5,0,col="black");
        panel.abline(0.1,0,col="black");
    },
    legend = list(
        inside = list(corner=c(1,1),x=0.98, y=0.98,
            fun="draw.key",
            args = list(
                key = list(x=0.99,y=0.94,corner=c(1,1),
transparent=F,background="white",fontfamily="Calibri",
                cex =0.7, border=0, columns=2, between=0.4,between.columns=0.1,
                lines=Rows(spl, 1:length(unique(newD$Pollutant))),
                text=list(keyTxt)
            )),
        inside = list(corner=c(1,1),x=0.34, y=0.98,
            fun="draw.key",
            args = list(
                key = list(
                    transparent=F,background="white",fontfamily="Calibri", cex =0.7,
border=0,columns=1,between=0.4,
                    points=Rows(sps, 1:length(unique(pointsD$Pollutant))),
                    text=list(levels(pointsD$Pollutant))
                )),
        ))))
    )
lfits;

# All data on separate plots with separate scales grouped by identifier
allPollutantsFree <- xyplot(
    Concentration ~ Distance| Pollutant,
    dataSet,
    as.table=TRUE,
    layout=(c(3,7,1)),
    #skip=c(rep(F,20),T,rep(F,3)),
    groups=ID,
    panel= function(x,y,...) {
        panel.xyplot(x,y,...);
    },
    xlab=list("Distance from edge
(m)",fontfamily="Calibri",fontface=2,cex=1.0),
    ylab=list("Concentration (units vary, but are consistent within
pollutants)",fontfamily="Calibri",fontface=2,cex=1.0),
    prepanel = function(x, y, ...) {
        list(xlim = range(x), ylim = range(0, y), dx = NULL, dy = NULL)
    },
    scales = list(relation = "free",fontfamily="Calibri"),
    par.strip.text=list(fontfamily="Calibri",fontface=2,cex=0.65)
)

```



```

allPollutantsFree;

#####
## Filename: Histograms.h
## Script purpose: Create histograms and barcharts to summarize the data
#####

## Generate indices for each distance increment of 50m
## *****
distCut <-
cut(dataSet$Distance,c(0,50,100,150,200,250,300,350,400,450,500,2000),right=F
)

distanceBins <- histogram(
  ~distCut,
  type="count",
  col="#A6CEE3",
  aspect=0.9,
  drop.unused.levels=FALSE,
  horizontal=FALSE,
  scales=list(x=list(tck=0,rot=45,at=1:11,labels=as.character(c("0-50","50-
100","100-150","150-200","200-250","250-300","300-350","350-400","400-
450","450-500",">500"))),fontfamily="Calibri",fontface=2),
  xlab=list("Distance bin (m)",fontfamily="Calibri",fontface=2),
  ylab=list("Number of distance/concentration
pairs",fontfamily="Calibri",fontface=2),
  #panel = function(x,y,...) {
  # panel.barchart(order(x),y)
  #}
  )
distanceBins;

## Which pollutants are studied?
dataSetI <- dataSet
dataSetI$Pollutant <- dataSetI$Pollutant[,drop=TRUE]

## YZ2002 and TK2005 both have unique papers with the same first six digit
identifier. Add new dummy factors for them, and rename so that they're
counted separately.
levels(dataSetI$ID) <- c(levels(dataSetI$ID),"TK2004","YZ2001");
dataSetI[dataSetI$ID=="TK2005b",]$ID <- "TK2004"
dataSetI[dataSetI$ID=="YZ2002b",]$ID <- "YZ2001"
dataSetI$ID <- substring(dataSetI$ID,1,6)
dataSetI$ID <- dataSetI$ID[,drop=TRUE]
row.names(dataSetI) <- 1:nrow(dataSetI)

dataSetI <- cbind(dataSetI,data.frame(Unique=rep(0,nrow(dataSetI))))
for(i in 1:nrow(dataSetI)) {
  currPol <- as.character(dataSetI[i,"Pollutant"])
  nextPol <- as.character(dataSetI[i+1,"Pollutant"])
  currID <- as.character(dataSetI[i,"ID"])
  nextID <- as.character(dataSetI[i+1,"ID"])
  if((currID != nextID) | (currPol != nextPol))
    dataSetI$Unique[i] <- 1
}
dataSetI$Unique[572] <- 1

```

```

dsnI <- dataSetI$Unique == 1
dataSetN <- dataSetI[dsnI, ]$Pollutant
pollutantCounts <- table(dataSetN)
pollutantCounts <- as.data.frame(pollutantCounts)
names(pollutantCounts) <- c("Pollutant","Freq")
row.names(pollutantCounts) <- 1:18

studiesCh <- barchart(
  Freq~reorder(Pollutant,-Freq),data=pollutantCounts,
  aspect=0.7,
  col="#A6CEE3",
  ylab=list("Number of studies",fontfamily="Calibri"),
  scales=list(x=list(rot=45),fontfamily="Calibri",fontface=2),
  )
studiesCh;

```

## Appendix B: Review of the near road pollution concentration literature

This review is structured roughly chronologically and includes all of the studies which are used in the analyses presented in the main text. The chronology is broken if a particular group of studies was completed as a series, and other studies are included as necessary to provide additional context or to compare and contrast methods and results. Each description begins with the roadway name and location, and the dates that measurements were completed are in underline and italics above the description of the study. Material is provided for more than 30 publications.

Early published work identifying near roadway areas as a concern for air quality came with a group of studies in the 1960s and 1970s. These were concerned with trace metal accumulation in soils and vegetation surrounding roadways. While some focused on lead (Cannon and Bowles 1962) others examined a suite of metals including cadmium, nickel, and zinc (Lagerwerff and Specht 1970), all known to be related to vehicular activity. They found increased contamination of soils near highways and intersections which decreased exponentially with distance.

An interest in the nature of air pollution concentrations in the same zone soon followed to complement the work on trace metals, emerging in the mid-1970s through work completed for CARB (Cahill et al. 1974) and the EPA. The difficulty of acquiring these early studies prevented their inclusion at the time of this writing.

### 1. Interstate 405 near Wilshire Blvd., Los Angeles, CA, 1978

Rodes and Holland (1981) sought to determine the influence of heavily traveled freeways on ozone and NO<sub>x</sub> concentrations to determine standards for siting regional-scale pollution monitors. Determining the extent of the roadway influence was critical for siting ozone detectors. Since NO<sub>x</sub> titrates ozone, detectors placed near roadways are not representative of downwind exposures. The authors studied the San Diego freeway in Los Angeles (200,000 veh/d) at six downwind locations (8, 32, 123, 198, 289, and 288 m from the freeway edge) and one upwind sampling location. Accounting for wind speed and direction, traffic counts and speeds, as well as ozone concentration during sampling, they characterized near road concentrations. They found

that a simple exponential relationship was able to describe the decrease in NO<sub>x</sub> concentrations as distance from the road increased and that the observed NO<sub>x</sub> concentrations changed with high, medium and low background ozone concentrations. Gradients for NO and total NO<sub>x</sub> were steep, decreasing by approximately 90 percent and 80 percent from their 8 m value by 123 m, respectively. Ozone levels were low and unchanged until 32 m, then increased to 123 m where leveling off occurred. The authors' final recommendations were to site NO<sub>2</sub> and ozone monitors at least 500 m downwind of a freeway which holds the roadway influence to less than 0.01 ppm.

2. *Kannana-dori Ave. and Route 17, north of Tokyo, Japan, 1982 and Omekaido Ave., Western suburbs of Tokyo, 1983*

Nitta et al. (1993), utilizing data from 1982 and 1983 for NO<sub>2</sub> and NO concentrations, attempted to quantify the effect of lightly traveled roadways (30,000 – 53,000 veh/d) in the suburbs of Tokyo, Japan on respiratory health. They also provide their emissions measurements. In general, concentrations immediately adjacent to the road were elevated above those further away. The gradient from 20 – 150 was more substantial for NO than for NO<sub>2</sub>, but no background concentrations or meteorology data were given.

3. *Arnhem and Wageningen, the Netherlands, 1994*

In the Netherlands, Janssen et al. (1997) conducted roadside measurements on lightly (8,900 veh/d) and moderately (15,000 veh/d) traveled roadways in the cities of Wageningen (pop. 35,000) and Arnhem (pop. 130,000). One data point located 0.5 m from the edge of road was taken and compared to background levels for PM<sub>10</sub>, PM<sub>2.5</sub>, and black smoke. PM<sub>10</sub> and PM<sub>2.5</sub> concentrations were on average 30 percent higher at the street than background while black smoke was 160 percent higher. Using elemental carbon concentrations estimated from black smoke measurements showed that the difference between street and background concentrations of PM could be attributed to elemental carbon. Since the Netherlands has a high proportion of private diesel vehicles the near road increment can be attributed largely to elemental carbon emissions from diesel fuel. A similar study was conducted in Amsterdam by Roemer and Wijnen (2001) who found that 24 hour averaged mass concentrations adjacent to the Einsteinweg (a 4-5 lane freeway) for PM<sub>10</sub> and PM<sub>1.0</sub> were substantially less elevated above background (30 percent) than was black smoke (300 percent).

4. A13 Motorway, Delft and Rotterdam-Overschie, the Netherlands, 1995

Roorda-Knape et al. (1998) conducted near road monitoring as a companion to health studies by Bunekreef et al. (1997) and van Vliet et al. (1997). All were conducted in the province of South Holland. The authors sought to validate the use of distance from road as a proxy for air pollution exposure, and coupled near road measurements of  $PM_{10}$ ,  $PM_{2.5}$ , black smoke, benzene, and  $NO_2$ , with indoor observations of  $PM_{10}$ , and  $NO_2$  at schools that participated in the health study. The reflectance of  $PM_{10}$  filters was also measured at these schools as a surrogate for black smoke. Six residential areas were selected with high proportions of homes within 300m of roadways. For logistical reasons, only two of the sites (Delft and Rotterdam-Overschie) could have detailed air quality monitoring completed, and both were located at different locations along the same freeway approximately 10 km from each other. These are the two contained in the database. At Delft, sites were located at 15, 115, 165, and 305 m from the A13 motorway (average daily traffic = 120,000) while Rotterdam-Overschie had sites located at 32, 82, 133, and 260 m from the A13. More extensive measurements of  $NO_2$  were taken to assess the variation in parallel to the road as opposed to only the horizontal gradient.

Averaging times varied by pollutant (continuous for benzene and  $NO_2$ , 15 min/hr for PM and black smoke). Black smoke and  $NO_2$  concentrations were high close to the motorway and declined with distance, showing a curvilinear gradient which was more pronounced during periods of high exposure (> 33 percent of time downwind). No such gradient was found for  $PM_{10}$ ,  $PM_{2.5}$  or benzene. Indoor concentrations of black smoke and  $NO_2$  in schools were less variable across different rooms in the same schools than were  $PM_{10}$  concentrations. Black smoke was correlated with truck traffic fraction and percentage of time downwind while  $NO_2$  was correlated with traffic intensity, percentage of time downwind, and distance of the school from the freeway. The authors did not measure background concentrations in any of the cases, so no inferences can be made about the increments caused by the road above this level.

5. A38 Motorway, Birmingham, United Kingdom, 1996-1997

Results comparable to those found by Roorda-Knape et al. (1998) were obtained in the United Kingdom by Shi et al. (1999) who examined the A38 motorway (30,000 veh/d) in Birmingham (pop.  $10^6$ ). A goal of the study was to determine the comparability of particle size

distributions measured with an electrical low pressure impactor and an SMPS. They conducted measurements at the roadside (2 m from the edge), 32 m from the edge, and at two rooftop urban background sites. They found number concentrations of ultrafine particles (in the size range 9.6 - 352 nm) elevated nine times above background levels, while PM<sub>10</sub> was elevated only by a factor of 1.5.

6. A38 Motorway, Birmingham, United Kingdom, 2000-2001

Harrison et al. (2003) attempted to quantify the contribution of various sources to the generation of roadside PM through analysis of the size distributions and chemical compositions of the particles. As lead has been phased out of gasoline, there has been a desire to identify other compounds which may be used as tracers of the contribution of road traffic.

Data was only taken at one distance from the road, and background concentrations were not measured as part of the study, but some data was used from a background monitoring site for comparison and was recorded in our database. PM<sub>10</sub> was measured and is recorded in our database, but other PM size fractions included 2, 1, 0.2 and were not recorded. Polycyclic aromatic hydrocarbons (PAH) of various types were also measured but were not recorded since only a single point was taken, and no background measurements accompany the observations.

Particle number (measured using a CPC) and NO<sub>x</sub> were used as indicators of traffic-related pollution. Particle number concentration was highly correlated with traffic, showing a similar diurnal pattern, and NO<sub>x</sub> had been used similarly in previous work by the authors. Additionally, CPC and NO<sub>x</sub> data were strongly correlated among themselves ( $r = 0.84$ ). Other metals showing strong correlations with CPC number were copper, molybdenum, barium, lead, and iron. Subsequent regression analyses and a principal components analysis led to inconclusive results; however, smaller  $R^2$ 's were noticed for many of the regressions (metals as the dependent variable, PM and CPC or NO<sub>x</sub> as the independent variables) as PM size fraction modeled decreased. This indicates a potentially greater contribution to trace metals in the roadside environment from resuspension processes than vehicle exhaust. Of the PAHs measured, methylphenanthrenes were most highly correlated with the traffic-related indicators, which are known to be markers of diesel exhaust.

7. I-80 Westbound near Davis, CA and Florin Rd. & Stockton Blvd., Sacramento, CA, 1994

The likely effect of roadway projects on the generation of  $PM_{10}$  “hotspots” —areas where air quality is particularly poor due to the contribution of specific sources in northern California was investigated by Cahill et al. (1994). They simultaneously measured traffic counts, meteorology, and pollution concentrations at both upwind and downwind locations for a major freeway and an urban intersection. A “sliding box” model was used to generate emissions factors per vehicle mile of travel at the three sites. An additional goal was a determination of the suitability of EPA’s standard AP-42  $PM_{10}$  emission factors for use on roadways in California. They determined that the intersection location had much higher potential to become a hotspot for  $PM_{10}$  generation than the freeway, due to increased amounts of material on the roadway surface. The AP-42 factors were found to overestimate the measured values. Experimental difficulties were experienced at the urban site, with measured mass from the particulate filters unable to reconcile with the aerosol composition. A follow up study was thus conducted by Ashbaugh et al. (1996). They generally found lower emissions factors than the previous study for the intersection location, and that its effects did not extend a great distance downwind with elevated  $PM_{10}$  concentrations falling to nearly background less than 100 m downwind.  $PM_{2.5}$  concentrations showed no elevation near the intersection nor did they vary diurnally.

8. Sunrise Blvd. & Greenback Ln., Sacramento, CA, 1997

A final report in this series, by Ashbaugh et al. (1998) examined the effect of winter weather conditions on hot spot generation at a heavily trafficked intersection (average 4710 veh/hr during study). The results indicated that even during periods of near-zero wind, ambient air quality standards for  $PM_{10}$  and  $PM_{2.5}$  were not exceeded near the intersection. Although measurements were taken at several distances from the road, the distances were not stated. A primary goal of the study was to quantify mixing behavior at various heights above the intersection, to see if hot spots would develop there. Accordingly, measurements were taken at 1, 3, and 9 m above the intersection.

9. Southeast Freeway, Brisbane, Australia, 1996

A study in Brisbane, Australia measured the horizontal profile of particle size distributions measured using an SMPS at two locations adjacent to the same freeway (Morawska

et al. 1999). “Site 1” was influenced primarily by freeway traffic, while “Site 2” was positioned near the confluence of many secondary roads and an interchange to the freeway. Vertical concentration profiles at three office buildings were also taken. In general at Site 1, total particle number concentrations (in the size range 16 – 626 nm, measured by the SMPS) were elevated 10 m from the freeway at Site 1, but showed no trend from 60 – 210 m. Site 2 showed an increased difference between the site closest to the freeway (27 m) and the others (43 and 73 m) relative to Site 1, but still no decreasing trend after the first site. There was a steep cliff at the end of Site 2, however, which could have influenced the results. The presence of an interchange also likely contributed to the increased particle concentrations observed at this site. No traffic volume or truck fraction data were given for the freeway. Analysis of the size distributions of particles indicated the same source both closest to and farthest from the freeway at both sites.

#### *10. Gateway Motorway, Tingalpa, Queensland, Australia, dates not specified*

A later study by the same group as Morawska et al. (1999) also investigated particle concentration and size distribution near a freeway in Tingalpa, Australia (Hitchins et al. 2000). Two size ranges of particle concentrations were measured, one using an SMPS (15 – 697 nm) and the other using an aerodynamic particle sizer (500 – 20,000 nm). PM<sub>2.5</sub> concentrations were also estimated using a DustTrak aerosol monitor. With wind from the road towards the monitoring stations, at a distance of 150 m, total particle number concentration reached 50 percent of its near road (15 m) value. Under parallel winds, the decrease to half occurred sooner, within 50 – 100 m. PM<sub>2.5</sub> mass concentration trended similarly, but only decreased about 25 percent over the entire measured distance (approximately 400 m).

No background concentrations were measured as part of the study, but observations were presented for periods where wind was towards the road and thus observations showed no roadway influence. These were recorded in our database as background values, but were not given for all wind conditions and may overestimate elevations above background when combined with the measurements for low wind speeds.

#### *11. Nijmegen, the Netherlands, 1999*

As part of a measurement campaign monitoring on-road PM mass and number concentrations using a mobile measurement unit, Weijers et al. (2004) performed a series of near



road measurements. During one day of their study, the measurement unit was driven away from unspecified heavily traveled roadways ( $> 10,000$  veh/d) and onto a quiet side street where it sampled for 30 seconds. Measurements were obtained for  $PM_{10}$ ,  $PM_{2.5}$ , and particle number concentration. Particle number concentration and  $PM_{2.5}$  (as total suspended particulates) were recorded in our database. The measured values from “residential areas” in Nijmegen were used as the background concentration. While the results showed concentration gradients near the roadway of similar shape for all PM mass measurements and particle number, the increment above background was far lower for PM mass than for particle number. The concentration of  $PM_{2.5}$  measured at 10 m from the road was 20 percent greater than background, while the ultrafine particle number was elevated 360 percent above background at 13 m from the road.  $PM_{2.5}$  decreased by 90 percent by 75 m,  $PM_{10}$  decreased 90 percent by 120 m, and the distance corresponding to a 90 percent drop in particle numbers was 210 m.

#### 12. VT5 Motorway and Savilahdentie, Kuopio, Finland, 1999

Tiitta et al. (2002) were motivated by a desire to properly site air quality monitoring stations and to produce data capable of validating roadside dispersion models for  $PM_{2.5}$ . The monitoring location for their study was located directly adjacent to an arterial road (Savilahdentie) but in close proximity ( $\sim 500$  m, p. 4058) to a freeway (VT5 motorway). Traffic ranges provided capture the range in flows expected on both road (low corresponds to arterial, high to freeway). Real-time traffic and meteorological (wind speed/direction, relative humidity, temperature) data were taken along with PM readings.

Noting that  $PM_{2.5}$  sources are varied, their model of total  $PM_{2.5}$  was broken into six separate terms, comprising both mobile (exhaust, non-exhaust) and stationary sources, urban and long-range transported background concentrations and resuspension processes not related to traffic flow. Mobile sources were further disaggregated by vehicle/fuel type and observational data was used as model input. The observed vehicular emissions were used then used to calibrate their model using multiple regressions. Additional data was used for the long-range transported and stationary components. Long range transport, vehicular (primary and non-exhaust) emissions, and other, accounted for 41, 33, and 26 percent of observed  $PM_{2.5}$  concentrations, although an application of the model to an independent dataset would be required for verification purposes.

The collected data was reported as well: traffic emissions caused an increase in  $PM_{2.5}$  concentrations of 30 percent from the furthest to the closest distances from road. This agreed well with model predictions ( $33 \pm 6$  percent).

*13. Hunts Point, Bronx, New York, NY, 1999*

Lena et al. (2002) conducted measurements of  $PM_{2.5}$  and elemental carbon at six intersections in the Hunts Point region of the Bronx which is heavily impacted by diesel truck traffic. (The sixth location an enclosed garden and will be reported as the background concentration on each of the intersection measurements). Wind direction and meteorology were not stated in the paper. Traffic volumes reported (both for cars and trucks) showed variability. Results reported here are representative means. They determined a relationship between truck density and elemental carbon concentrations ( $R^2 = 0.84$ ) but that between truck density and  $PM_{2.5}$  was much weaker ( $R^2 = 0.52$ ). This highlights the importance of mobile sources to elemental carbon and regional sources for  $PM_{2.5}$ .

*14. Avenida da Amizade Road, Rua do Campo Road, Rua de Ferreira do Amaral Road, and Rua da Ribeira do Patane Road, Macao, China, 1999*

Wu et al. (2003) investigated the chemical composition of  $PM_{10}$  and  $PM_{2.5}$ , providing point-roadside data was given and was recorded in our database. Distances from road were only reported as 1 – 2 m, therefore, 1.5 m was recorded in our database as a point estimate.

Organic carbon was determined to be the largest component of both roadside and background concentrations of PM. However, the study did not attempt to account for sampling artifacts (adsorption of gaseous organic compounds, or evaporative loss of volatile organic compounds) so the organic carbon data was not recorded in our database. Crustal materials exhibited high concentrations both near road and at background, indicating a low traffic-related component. Elemental carbon was a major near road component and was found to be higher next to roads with a higher proportion of diesel traffic, supporting its use as a diesel tracer. Both organic carbon and elemental carbon were reported for  $PM_{10}$  and  $PM_{2.5}$ . Elemental carbon is known to reside mostly in the submicrometer range, and background values were given for both.

*15. Avenue Leclerc & Place Basch, Paris, France, 2001*

Vardoulakis et al. (2005) conducted this study in a highly urbanized area of Paris, France, with a mix of traditional six-storey buildings, modern high rises, and some detached homes. A week of “intensive” sampling was completed comprising vertical measurements taken along the sides of buildings, in addition to the roadside monitors at the study intersection (66,000 veh/d). Trees were found to disrupt the transport of benzene upwards beyond the second floor of apartment buildings.

Ten roadside monitors were placed along Avenue Leclerc to detect small spatial scale variation in pollution concentrations. Each monitor was placed exactly at the roadside or within the traffic stream (on medians). Strong correlations were found between all measured volatile organic compounds (VOCs) and CO, and regression analysis was used to model benzene concentrations based on CO. The authors found that wind speed and direction had a large effect on observed concentrations—higher concentrations of pollution were observed on the leeward side of the street. The highest benzene concentrations were observed in October during low-wind speeds, and the lowest in August due to low traffic density, again illustrating the complex interplay between the factors contributing to observed concentrations. Overall seasonal trends were not visible. High variability was found due to differences in urban form and meteorology, leading to the realization that the siting of monitoring equipment is crucial. Measurements conducted where the intake of undiluted exhaust gas is possible are not representative of exposure. Dispersion models were tested for consistency with observations and were found to overestimate concentrations in all cases but this could be due to the area's complex geometry.

Concentration results were only presented for benzene. Other data was used to populate correlation matrices, to complete linear regressions, or was presented in high time resolution, making the extraction of an average for database input excessively difficult.

*16. Interstates 580 and 880, State route 92, East Bay, San Francisco, CA, 2001-2002*

The East Bay Children's Health Study was designed to assess the links between exposure to traffic related exhaust and respiratory illness in children attending 10 schools in the East Bay of the San Francisco Bay Area, which has generally good regional air quality. The schools were selected to represent locations upwind and varying distances downwind from the freeways under study, and neighborhood measurements were taken at three of the schools to assess outdoor

concentrations at childrens' residences. Residences were found to be good indicators of school exposure for those upwind or far downwind, but near downwind homes underestimated or overestimated on the basis of distance from road.

Responses to questionnaires ( $n = 1,109$ ) on health outcomes (asthma, bronchitis, and chest illness) were modeled via logistic regression using air pollution ( $PM_{10}$ ,  $PM_{2.5}$ , black carbon,  $NO_X$ ,  $NO_2$ ,  $NO$ ) monitoring data taken at each of the schools as independent variables. Since most students resided in the surrounding neighborhoods, school exposure was taken as a surrogate for personal exposure.

Kim et al. (2004) reported the epidemiological results. In general,  $NO_X$ ,  $NO$ , and black carbon showed stronger associations with respiratory illness, gauged by increasing odds ratios, than  $PM_{10}$ ,  $PM_{2.5}$ , and  $NO_2$ .  $NO_X$ ,  $NO$ , and black carbon were also highly correlated. An additional model using distance downwind from a major road instead of pollution concentrations showed similar or stronger odds ratios, indicating that the roadway is a key source of exposure to these pollutants.

In general black carbon,  $NO_X$ ,  $NO$ , and  $NO_2$  to a lesser extent increased in concentration as the schools decreased in distance to a major roadway (either I-580 or I-880) within the range of 300 m. The authors reported concentrations taken at these schools and averaged over periods of tens of weeks. Reported concentration values therefore do not all originate from the same location on the same freeway. Each school was identified as being close to a particular freeway and thus was exposed to the annual average daily traffic associated with that freeway. Schools upwind or far downwind (schools 1, 3, 4, 6, and 7 from the study) did not have statistically distinguishable differences in pollutant concentrations. Background concentration values reported in the paper therefore reflect average for schools greater than 300 m from a major traffic source. Singer et al. (2004) showed that concentrations of  $NO_X$  and  $NO_2$  measured at those schools were similar to those taken at the closest stationary monitoring station—within 350 m upwind, concentrations were similar to regional monitoring station data and from schools located >1000 m downwind of the freeway.

Singer et al. (2004) took hourly wind speed, direction, and temperature were taken primarily to validate classifications by wind direction. Sites to the east or northeast of freeways were predominantly downwind from traffic emissions although wind patterns shifted during the winter months. Near road effects were somewhat higher in spring when the school was directly

downwind of the freeway for almost the entire day and somewhat lower in fall, when nighttime winds shifted.

*17. E4 Motorway, 28 km north of Stockholm, Sweden, 2003*

Gidhagen et al. (2004) attempted to quantify the geographical and temporal variations of ultrafine particles to better model exposure levels for urban populations. By investigating a rural highway they expected low particle surface area (rapid dilution) and high vehicle speeds. These conditions favor the generation of ultrafines. The effect of temperature was also investigated, and aerosol dynamics were explicitly accounted for. The authors performed near road monitoring at 10 m and 63 m from the E4 motorway (52,300 ADT). The results showed that total NO<sub>x</sub> and total particle number were highly correlated while the ratio of particle number to NO<sub>x</sub> decreased as temperature increased.

Results suggest that, for assessing urban exposure, deriving emissions factors from roadside monitors between 10 – 15 m from the edge of road would be adequate data for Gaussian line modeling to assess exposure to total particles.

*18. Interstates 405 and 710, Los Angeles, CA, 2001-2005*

A series of studies conducted by researchers at UCLA and USC became heavily cited in the subsequent study of pollution concentrations near heavily traveled roadways. Zhu et al. (2002b) examined CO, black carbon, and particle mass and number concentrations at five locations downwind of the I-405 freeway (13,900 veh/hr during peak, <5 percent trucks) in Los Angeles, California. They found that with increasing distance, ultrafine particle size distributions shifted to greater diameters and number concentrations dropped dramatically. Number concentration decreased to approximately half of its value at 30 m between 90 m and 150 m and reached background by 300 m. CO and black carbon tracked particle number well, leveling off by 150 m from the freeway.

Zhu et al. (2002a) conducted similar measurements at the I-710 freeway which has a large proportion of diesel truck traffic since it functions as a goods movement corridor for the ports of Los Angeles and Long Beach (12,000 veh/hr during peak, 25 percent trucks). Total particle number concentration (6 – 25 nm) decreased 80 percent in first 100 m before leveling off somewhat after 150 m and reaching background at 300 m.

Particle number, black carbon, and CO all exhibited a similar exponential decay pattern at both the 405 and 710 freeways while absolute concentrations of black carbon and CO differed, as would be expected given the difference in fleet composition on each section. I-405 generally had twice the CO concentration and a third of the black carbon concentration as I-710.

An identical study was carried out by Zhu et al. (2004) during the winter of 2002 at the same freeway locations and same distances. They found a distinct mode in the 20 – 30 nm size range for the 710.

To examine the effects of meteorology on observed concentrations, Zhu et al. (2006) conducted air quality measurements over seven nighttime periods (11pm – 4am) and compared them to their previous results from the I-405 freeway in Los Angeles (Zhu et al. 2002b). Particle number concentrations decreased more slowly during the night than during the day, to reach 40 percent of their original value at 30 m by 300 m, but even by 500 m the background level had not been reached. This implies that the nighttime contribution from freeways to local air quality is more substantial than during the day because of increased local traffic during this time (the day) which raises background levels of pollution.

Although traffic volume was only 25 percent of the daytime value, particle number concentrations were approximately 80 percent of their daytime value. This was the contribution of meteorology. Lower temperatures correspond to higher rates of particle production per vehicle. Speeds are also higher at night, which is proportional to particle number concentrations. There were more trucks at night than during the day. However, the most important factor appeared to be the relative humidity which increases the number concentration of the nucleation mode. None or very weak  $PM_{10}$  and  $PM_{2.5}$  gradients were observed indicating that policies to limit either may be ineffective at reducing total particle number concentrations.

No trends were observed at upwind locations, indicating that using proxy indicators for traffic exposure may be inappropriate since actual AP exposure is highly dependent on wind and meteorological conditions that vary over the day and by season. The authors noted, “...many epidemiology studies have relied on distance from major roads or self-reported traffic density as a proxy for exposures. It is clear ... that none of these is a good measure of exposure to traffic exhaust” (Zhu et al. 2006, p. 2534).

*19. Interstate 710, Los Angeles, CA, 2006*

The I-710 was investigated again in the Winter of 2006 by Ntziachristos et al. (2007b) who sought to classify the physical and chemical properties of particles near the freeway in detail including particle number, surface area and mass. The authors used a single location, 10 m from the edge of the shoulder to measure both gaseous and particulate concentrations. Size distributions were measured using a SMPS (16 – 638 nm) and particles were counted with a CPC (>6 nm). Gaseous pollutants were also measured.

They found that particle number concentration near the freeway was overwhelmingly in the less than 16 nm size range. This comports with the results of Zhu et al. (2004) and results from a comparison of the SMPS and CPC numbers. Further, particle size distribution was bimodal, peaking in both the nucleation and accumulation modes which is indicative of roadways dominated by diesel vehicles.

Particle emissions dropped substantially on a day when the freeway was closed to traffic, but organic carbon concentrations did not drop as much nor did they correspond as closely to traffic volumes as elemental carbon and black carbon, indicating a secondary source and secondary formation mechanisms. All measures of particle intensity except one peaked at 4 pm when traffic was congested, strengthening the link between traffic and particle emissions. CPC number increased with time of day (11 am - 7 pm) because of the nucleation of small particles as temperatures dropped.

A further study goal was to investigate the appropriateness of using the dilution ratio (which compares the CO<sub>2</sub> ratio in exhaust gas to that at the sampling site) to decouple the influence of sampling site location, meteorology and fleet effects on measured values. This was accomplished by a comparison to two other studies.

Regression modeling was also undertaken, using surface area concentration and the particle volume in the nucleation and accumulation modes. Light and heavy organic carbon were associated with the nucleation mode and accumulation mode volume, respectively supporting previous evidence of a unique particle mode adjacent I-710 in the size range 20 – 30 nm. Further, predominantly gasoline fueled vehicles have shown an organic carbon profile shifted towards higher mass.

The deposition and composition of trace metals (including nickel, copper, chromium, lead, and zinc) in the vicinity of both freeways was also studied at approximately the same time.

Sabin et al. (2006) investigated the chemical composition of particles deposited at three distances from the Interstate 405, and one background location in 2003. The freeway appeared to be a source of particles greater than  $6\text{ }\mu\text{m}$  in diameter. Deposition of metals reached background quickly, somewhere between 10 and 150 m. A simple modeling effort based on Gaussian plume dispersion showed good agreement with empirical results. The study was limited by the small sample size.

20. State Route 110, Los Angeles, CA, 2005

Kuhn et al. (2005a) and Kuhn et al. (2005b) studied State Route 110, a freeway with heavy duty truck traffic prohibited in the summer of 2004 and winter of 2005, respectively. The studies were motivated by the knowledge that the organic constituents of PM emitted from vehicles are semivolatile and that observed concentrations and size distributions can thus be affected by meteorology as well as the chemical properties of the compounds. The authors sought to characterize how size distributions shift due to volatility and to draw conclusions for exposure to PM. Two locations adjacent to the roadway were used in each paper (results were reported at 2.5 and 150 m) and the chemical composition and mass of particles in the coarse ( $2.5 - 10\text{ }\mu\text{m}$ ), accumulation ( $0.18 - 2.5\text{ }\mu\text{m}$ ), and ultrafine ( $< 0.18\text{ }\mu\text{m}$ ) modes were measured between 12 pm and 7 pm on weekdays.

The results showed that higher wind speeds and drier conditions in the summer led to higher mass concentrations in the coarse mode than in winter. The accumulation mode masses were similar at both measurement locations, indicating that that size fraction represented an urban background aerosol. Concentrations broken out by species showed little variation between the two locations in winter except for organic carbon, elemental carbon, Cu, and Ba. However, organic carbon was affected by a sampling artifact in which gaseous organic material was absorbed by the measurement filter. In general, concentrations of all species were higher in the summer except for organic carbon and elemental carbon which are known to track traffic emissions. They are thus affected by atmospheric inversion caused by colder temperatures.

Observed concentrations did not include a background value, and so are not extensive enough to determine when that value is reached. Recorded in our database were values of  $\text{PM}_{10}$ ,  $\text{PM}_{2.5}$ , elemental carbon, and sulfur.



### 21. Interstate 80 near Sacramento, CA, 2000-2001

Nanzetta and Holmen (2004) also attempted to quantify the effects of meteorology and traffic volume on observed particle number concentrations. They investigated a section of I-80 near Sacramento, California (6330 veh/h during the study period with 10 percent heavy trucks) in both fall and summer months. A single monitor was placed downwind at 3 m from the edge of road and an upwind monitor was placed at 6 m from the roadway edge in winter and 9 m from the edge in summer. Ratios between downwind and background observations ranged from 14.6 in winter to 1.3 in summer in the size range 6 – 237 nm, and 6.5 in winter to 2.1 in summer in the size range 23 – 898 nm. These differences are due largely to temperature and relative humidity, since nucleation of new particles decreases as temperatures increase and relative humidity decreases.

Statistically significant relationships between observed concentrations and roadway parameters and meteorology were difficult to ascertain in many cases. While relationships were found with heavy duty vehicles and particle numbers in summer, no relationship with any vehicle types was found in winter. Temperature, relative humidity, and solar radiation were significantly correlated with upwind concentrations, while only temperature and wind speed showed significant correlations with downwind concentrations. Their study again highlights the importance of daily and seasonal trends in assessing observed pollutant concentrations and points to a need for better time resolved traffic and concentration measurements.

### 22. Interstates 71 and 75, Cincinnati, OH, dates not specified

Reponen et al. (2003) studied the I-71 and I-75 freeways (135,000-150,000 veh/d, 10 percent trucks) in Cincinnati, Ohio, measuring particle number, elemental and organic carbon, and PM<sub>2.5</sub> mass concentrations. The measurements were conducted in support of a larger epidemiological effort, the Cincinnati Childhood Allergy and Air Pollution Study. Condensation nucleus counters, measuring in the size range 0.02-1 µm, showed a steep gradient in ultrafine particle concentrations when moving away from the freeway. Concentrations decreased by 50 percent between 50 and 150 m. Background concentrations were measured at 1600 m downwind from the freeway and were reached by 800 m. A minimal gradient was observed for coarse particles as measured by an optical particle counter (size range 0.3-20 µm). In general, however,

“traffic particles could not be distinguished from the background levels beyond 400 m from the highway” (p. 562).

Exposure to traffic pollution was detected at distances up to 400 m using sulfur (believed to be a surrogate for diesel exhaust) as an indicator although results were somewhat inconsistent (and not discussed) between sampling periods. Elemental carbon, another indicator of diesel exhaust, exhibited elevated concentrations up to 400 m downwind as well.

### 23. Stockholm County, Sweden; Munich, Germany; and Rotterdam, the Netherlands, 1999

Hoek et al. (2002) sought to quantify the average annual concentration and spatial variability of  $PM_{2.5}$  in Munich, Stockholm and the Netherlands. Approximately 40 sites were selected in each area and were broken down into background (urban/rural) and traffic. Urban background was further disaggregated by proximity to city center. Results are reported as means and ranges for each type of site.

The authors found more spatial variability for  $PM_{2.5}$  than previous literature (17 to 18 percent higher at traffic sites than at urban background locations), but the  $PM_{2.5}$  absorption coefficient exhibited even larger variability (31 to 55 percent increase at traffic sites versus urban background concentrations), supporting previous findings suggesting that measuring only  $PM_{2.5}$  (or  $PM_{10}$ ) concentrations as a traffic-related air pollution indicator can underestimate the contrast between background and traffic-related emissions exposure. Detailed data on distance from road was not given, but mean distance from road for all sampling sites within each of the three areas ranged from 112 to 139 m. In Stockholm and the Netherlands the maximum distance from road was 450 m while for Munich it was 930 m.

The sampling approach utilized four two-week mean samples at each site taken over a period of a year and calibrated using a monitor in continuous operation at one particular site in the study area. The authors suggest that similar approaches could be used in other cities to assess spatial variations in pollution that could contribute to differential epidemiological risks.

Using similar methods, a later paper by the same group (Cyrus et al. 2003) investigated correlations between annual averages of elemental carbon,  $PM_{2.5}$ ,  $PM_{10}$ , and PM absorption coefficients in the three study areas to determine the extent to which absorption coefficient of PM filters can be used as a surrogate for elemental carbon concentrations. They conducted direct measurements of elemental carbon using a two part process whereby organic carbon is first

removed by thermal desorption and the remaining elemental carbon is oxidized to CO<sub>2</sub> and water. They found that contrasts between traffic and background were greater for elemental carbon and PM<sub>2.5</sub> absorption coefficients than PM mass concentrations, and that the PM<sub>2.5</sub> absorption coefficient better tracks spatial variability in combustion derived particles than PM<sub>2.5</sub>. On net, their findings supported the use of PM filter absorbance as an inexpensive surrogate for elemental carbon concentrations.

#### 24. Highway 15, Montreal, Quebec, Canada, 2001-2003

A pair of studies conducted by Gilbert and colleagues at Health Canada and Environment Canada (Gilbert et al. 2003; Gilbert et al. 2005) sought to characterize the exposure profile of NO<sub>2</sub> adjacent roadways in Montreal, Quebec.

Gilbert et al. (2003) conducted measurements at the Decarie Highway in Montreal (185,000 veh/day) over a week in September, 2001. They found that the major NO<sub>2</sub> decrease occurred within 200m of the highway. An estimated regression model including the logarithm of distance from roadway and a binary variable for upwind/downwind location had an R<sup>2</sup> of 0.97. Downwind concentrations were significantly greater than upwind concentrations at similar distances, although no clear “background” value was provided so it is not clear when or whether NO<sub>2</sub> concentrations became uninfluenced by contributions from traffic. Here the decline was much more gradual at increasing distances and almost even between upwind/downwind samplers. This could have been due to the temporal nature of the study (values averaged over 7 days as opposed to just peak periods). The authors conclude that distance from a major road may adequately characterize exposure to NO<sub>2</sub>.

Gilbert et al. (2005) developed a multiple regression model to predict NO<sub>2</sub> concentrations based on land use characteristics. The results of 67 monitoring locations were included in the analysis. The final model included positive associations with traffic count on the nearest highway, length of highways within 100 m, the length of major roads within 100 m, length of minor roads within 500 m, population density within 2000 m, and negatively associated with the area of open space within 100 m and distance from the nearest highway. Final model R<sup>2</sup> was 0.545. The authors suspect that higher values would have been obtained if wind speed and industrial sources were better accounted for.

25. Metropolitan Highway and various urban arterials, Montreal, Quebec, Canada, dates not specified

Smargiassi et al. (2005) also conducted near road measurements in Montreal and sought to assess the spatial and temporal variability of elemental carbon (measured using the  $PM_{2.5}$  absorption coefficient),  $PM_{2.5}$ , and  $NO_2$ . They also assessed the effect of traffic density on the observed concentrations. A primary study goal was to draw conclusions regarding the exposure of individuals residing adjacent to urban roads to particulates.

Four sites were chosen with differing traffic intensity but free from stationary sources of pollution. The first represented background concentrations—a quiet residential street. Two and three were on urban arterials, while four was adjacent to the Metropolitan Highway. Sites were separated by approximately 1 km, would therefore show an influence only from their closest roadway. Thus, this study, instead of moving further from roads, held distance from road constant and increased the traffic volumes. All three traffic sites were located “less than 10 m from the curb.”

A gradient was observed with increasing traffic volume for  $PM_{2.5}$  absorbance and  $NO_2$ .  $PM_{2.5}$  was approximately equal across all sites, supporting the regional nature of  $PM_{2.5}$  emissions. Although results showed spatial gradients in the urban area studied, the authors were unable to draw definitive conclusions about personal exposure.

26. M50 Motorway, Dublin, Ireland, 2003

A pair of studies published by colleagues at Trinity College in Dublin, Ireland, examined the concentrations of 11 hydrocarbons at three distances downwind and upwind (six receptors total, each at 25, 120, and 240 m from the edge of road on both sides) of the primary beltway in Dublin, Ireland, the M50. The data collection and results are described in O’Donoghue and Broderick (2007) and the data is compared to dispersion modeling using CALINE4 and COPERT III in Broderick and O’Donoghue (2007). The studies were motivated by a lack of near road hydrocarbon measurements previously reported in the literature.

Consistent meteorology in terms of temperature and wind speeds persisted over the sampling campaign, which lasted five weeks and took place during the morning peak period (7 am – 10 am). Small modulations in wind speed (0.85 m/s average when from the east versus 1.28 m/s from the west) were enough to change the rate of decay to background concentrations for the

non-alkenes monitored (*iso*-pentane, *n*-pentane, ethane, propene, 1,3-butadiene, acetylene, and benzene). By 240 m from the road under all wind conditions, background concentrations were close to being reached. For the lower wind speeds, between 20 and 35 percent of the concentration at the nearest monitor remained by 240 m. If an edge concentration would have been interpolated, however, this number would have been smaller. Monitored alkenes (ethane, propane, *iso*-butane, and *n*-butane) did not vary with distance from road.

Benzene was grouped with other measurements of that pollutant into its own category, while the other measurements were grouped together as alkanes or non-alkane hydrocarbons in our database, since no other studies overlap with the non-Benzene compounds measured by these researchers.

27. Interstate-440, Raleigh, NC, 2006

Hagler et al. (2008) investigated the relationship between traffic and ultrafine particle number concentrations. They found variability in terms of absolute particle number measurements, but a similar pattern of decay on each observation day. They noted that benzene, toluene, NO, CO, black carbon correlated with the decay of ultrafines, but that secondary pollutants did not correlate (PM<sub>10</sub>, PM<sub>2.5</sub>, NO<sub>2</sub>). They did not quantitatively assess the influence of each of the parameters on observed concentrations.

28. Highway 71, Interstate 35, FM-973, Austin, TX, 2007

Clements et al. (2008) and Zhu et al. (2008) performed an extensive measurement campaign adjacent to Highway 71 (arterial), Interstate 35 (freeway), and a truck traffic-dominated freeway (FM-973). A specific study goal was to determine how the traffic composition and freeway design affects observed concentrations. Data collection methods mixed stationary monitoring units with a mobile platform that stopped at fixed distances from the roads.

Clements et al. (2008) reported those measurements from the arterial. In addition to ultrafine particles (< 100 nm), criteria pollutants (CO, NO<sub>x</sub>, PM<sub>2.5</sub>, VOCs) and several other organic compounds were measured—polyaromatic hydrocarbons, alkenes, hopanes, and several carbonyls. Traffic-related pollution generally decayed to background concentrations by 150 m. Some complex relationships with distance were noted for specific carbonyls (Acetaldehyde) and alkenes (C28) which increased with distance. These were not recorded in our database.

Suggested reasons include chemical reaction with other exhaust species. Other alkenes and carbonyls showed the expected decreasing relationship.

Zhu et al. (2008) reported the ultrafine particle number results observed at Highway 71, and the normalized results for all three roadways. The data were normalized to unit traffic intensity and wind speed, which is only possible if the primary data are accessible. Their results showed good agreement between roads when normalized ultrafine particles were plotted with distance, strengthening the importance of traffic intensity and wind speed as factors affecting observed near road concentrations.

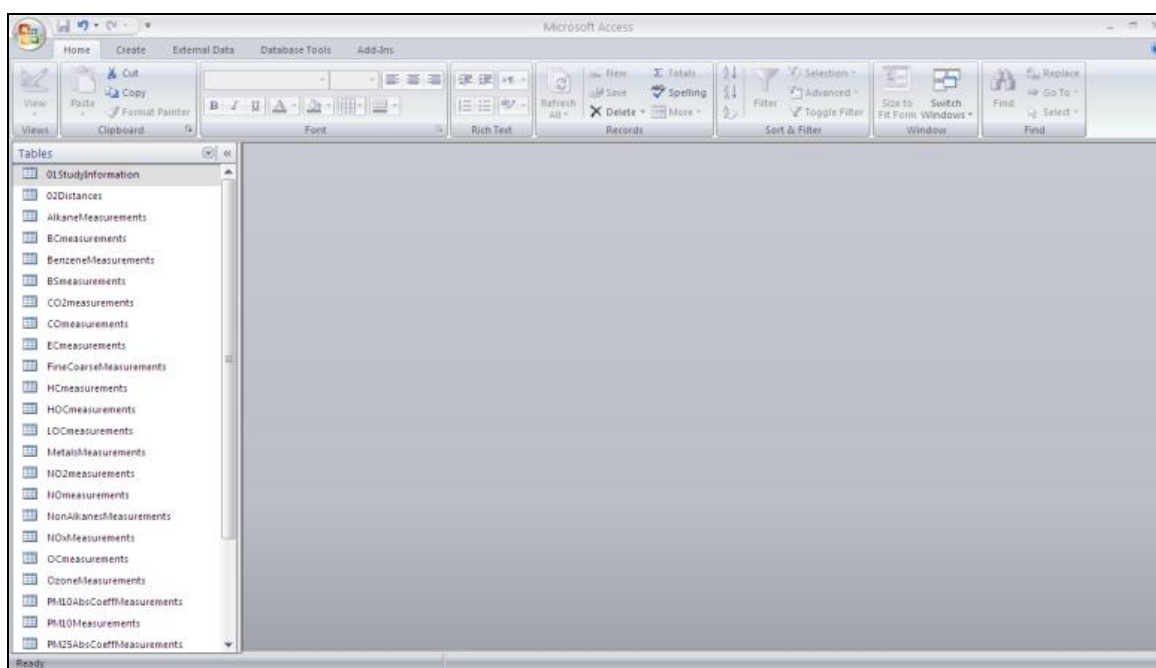
## Appendix C: Database

### List of figures and tables

Figure C 1: Concentration database opened in Microsoft Access. ....	105
Figure C 2: “01StudyInformation” opened in Microsoft Access.....	106
Figure C 3: Relationships in the database. ....	109
Figure C 4: “02Distances” opened in Microsoft Access. ....	110
Figure C 5: “PM25measurements” opened in Microsoft Access. ....	111
Table C 1: Explanation of Labels Used in the “01StudyInformation” Table in the Database ...	108

As discussed in the main text, a Microsoft Access database was compiled to store distance/concentration pairs and other relevant details from each study. This appendix contains an overview of that database.

Upon opening the database, Microsoft Access appears as shown in Figure C 1. The database is composed of many tables, all of which are listed alphabetically in the left hand pane of the opening screen. Unless minimized, the table listing remains docked in this location. The most important table in the database is entitled “01StudyInformation” and appears first in the list. It contains all of the non-distance/concentration data about the paper. Double clicking “01StudyInformation” opens the table, as shown in Figure C 2.



**Figure C 1: Concentration database opened in Microsoft Access.**

Each database table has two components: records and fields (also known as attributes). Records refer to rows of the table, and fields to the columns. In our database, each record represents one study, or a unique set of measurements from one study. For “01StudyInformation”, the meaning of each field is contained in Table C 1. The listed fields correspond to those shown along the top of the table “01StudyInformation” in Figure C 2. The identifier listed for the study is used throughout all of the tables to consistently identify the paper source of the information. Establishing this naming convention allows full use of the *relational* nature of the database whereby information in different database tables can be traced back to the



central record located in “01StudyInformation.” Using the unique identifier in subsequent tables to refer to the entry in “01StudyInformation” also upholds the integrity of the database by establishing explicit relationships between different elements of tables meant to refer to the same individual study. One cannot enter data elsewhere in the database on a study that does not already exist in “01StudyInformation,” for example.

The relationship structure of the pollution concentration database is shown in Figure C 3. In the first two tables (“01StudyInformation” and “02Distances”) the unique identifier is designated as the *primary key* for each record (indicated by the image of a key beside that field in those two tables). The primary key is a field that contains a unique entry for that record within one table. The links between tables shown in the figure indicate that one element of each table is matched to another field in another table. The link also indicates the relationship type which can be either *one to one* (indicated by a “1” at both ends of the link) or *one to many* (indicated by a “1” at one end and a “∞” at the other end of the link). In a one to one relationship, only one element of each record may appear in both related tables. In a one to many relationship, one record element appears in one of the tables, but it may appear many times in the other table. This means that there can only be one record per unique identifier within the first two tables, but that many records can use the same identifier in all of the other tables of the database.

Identifier	Unique	YearofPub	BriefTitle	Author	Location	StartDate	EndDate	DateNotes	Tr
AC2008	✓	2008	Air Pollutant Concentrations near TX roadways	Clements et al., 2008	Highway 71, Austin, TX	4/1/2007	6/1/2007	Different pollutants measured on different days	4
AS2003	✓	2003	...mobile source effect on curbside...	Sapkota and Buckley, 2003	Baltimore Harbor Tunnel Tollbooth Facility				
AS2005S1	✓	2005	Small-scale spatial variability... in Montreal: a pilot study	Smargiassi et al., 2005	Rue Molson (Background site), Montreal, Quebec	5/6/1900	6/28/1900	Year unknown, weekdays only	All
AS2005S2		2005	Small-scale spatial variability... in Montreal: a pilot study	Smargiassi et al., 2005	Rue D'Iberville, Montreal, Quebec	5/6/1900	6/28/1900	Year unknown, weekdays only	All
AS2005S3		2005	Small-scale spatial	Smargiassi	Bldv St-	5/6/1900	6/28/1900	Year unknown,	All

Figure C 2: “01StudyInformation” opened in Microsoft Access.

Distance/concentration data is contained in two tables: “02Distances” and another entitled “Xmeasurements” where “X” refers to a specific pollutant. “02Distances” contains data on the distances from road used by each study and is shown in Figure C 4. The unique identifier appears in the first column of the table followed by the distances at which pollution measurements were taken within that study (labeled “Dist1” through “Dist10”). The reference point (center line, curb, etc.) is included in a note. Once data on the roadway cross section was obtained via Google Maps, all distance data were normalized to an edge of road reference point. The normalized data appear in columns labeled “NDist1” through “NDist10.”

**Table C 1: Explanation of Labels Used in the “01StudyInformation” Table in the Database**

<b>Field</b>	<b>Explanation</b>
Identifier	Unique article identifier of the form: author's two initials, four digit year and an additional character if needed (e.g, AK2007a, DE2002b, etc.)
Unique	Checkbox indicating whether this is the first entry for this study
YearofPub	Publication year
BriefTitle	Article title, shortened if necessary (useful for identifying articles conclusively)
Author	Article citation in Author, Year format
Location	Location of study: at least city and facility
StartDate	Start date of study
EndDate	End date of study
DateNotes	Notes to clarify which dates were actually included in the study
TimeOfDay	Hours of the day during which the study was undertaken
Season	Season during which the study was undertaken
TrafficRange1	Low value of the traffic volume during the study
TrafficRange2	High value of the traffic volume during the study
TrafficRangeUnits	Units of the traffic volume used for the ranges
SpotMeanTrafficVolume	Mean traffic volume
TrafficVolumeUnits	Units of the mean traffic volume
PercentTruck	Percentage of heavy vehicles on the facility
AADT	Average annual daily traffic
FacilityType	Freeway, arterial, or residential
Shoulder	Checkbox indicating whether a shoulder was present
NumLanes	Number of lanes
Chemistry	Checkbox indicating whether a chemical analysis was undertaken on the constituents of particles
AveragingTime	The time period over which results were averaged
DataRecorded	Checkbox indicating whether the study included data appropriate for storage in the database
Notes	Summary findings

**Figure C 3: Relationships in the database.**

The tables labeled “XMeasurements” can be seen in the leftmost pane of Figure C 1 beginning with “AlkaneMeasurements” and “BCmeasurements.” As an example, the table for storing concentration measurements for PM<sub>2.5</sub> is shown in Figure C 5. Each record includes entries for the study identifier, and all concentration measurements for that pollutant given in the study labeled “Meas1” through “Meas10.” An additional column labeled “KeyCondition” indicates to which set of measurements that record refers, since any individual study may have taken more than one set of measurements for any pollutant. As mentioned in the main text, if a study included any disaggregate measurements, they were included as separate measurement sets. Finally, a dummy numeric field that increments by one with each new entry is included for each record. Since the study identifier was not unique in the “Xmeasurements” tables, another primary key was required. This was set to the numeric dummy field.

Identifier	Dist1	Dist2	Dist3	Dist4	Dist5	Dist6	Dist7	Dist8	Dist9	Dist10	DistanceMeasuredF	NDist1
AC2008	15	45	82	105	138	0	0	0	0	0	0	Edge of road
AS2005S1	10	0	0	0	0	0	0	0	0	0	0	Curb
AS2005S2	10	0	0	0	0	0	0	0	0	0	0	Curb
AS2005S3	10	0	0	0	0	0	0	0	0	0	0	Curb
AS2005S4	10	0	0	0	0	0	0	0	0	0	0	Curb
BS2004D	60	130	200	230	1200	1400	0	0	0	0	0	Edge of road
BS2004U	350	360	1500	1700	0	0	0	0	0	0	0	Upwind measures
CR1981	8	32	123	198	289	388	0	0	0	0	0	Edge of road
EW2004M	10	25	43	60	78	93	113	128	145	163	163	Edge of road
EW2004U	13	33	55	75	97.5	122.5	142.5	165	185	197.5	197.5	Edge of road
GG2004	25	45	57	70	82	107	132	155	307	0	0	Center line
GH2002	124	0	0	0	0	0	0	0	0	0	0	Mean of all study sites
GH2008	25	50	100	175	0	0	0	0	0	0	0	Unspecified, edge ass
HN1993B	0.1	20	50	150	0	0	0	0	0	0	0	Edge of road
HN1993C	0.1	20	150	0	0	0	0	0	0	0	0	Edge of road
JC2003G	10	0	0	0	0	0	0	0	0	0	0	0
JC2003N	6	0	0	0	0	0	0	0	0	0	0	0
JC2003S	19	0	0	0	0	0	0	0	0	0	0	0
JH2000a	15	55	95	135	215	295	375	0	0	0	0	Edge of road
JK2004	60	130	200	230	350	360	0	0	0	0	0	0
JS1999	2	32	0	0	0	0	0	0	0	0	0	Edge of road
LA1996	9	50	88.5	0	0	0	0	0	0	0	0	Corner of intersection
LG2004	10	63	0	0	0	0	0	0	0	0	0	Edge of road
LM1999A	10	60	80	100	160	210	0	0	0	0	0	Edge of road
LM1999B	27	43	73	0	0	0	0	0	0	0	0	Edge of road
LN2007	10	0	0	0	0	0	0	0	0	0	0	Shoulder
LS2006	10	150	450	0	0	0	0	0	0	0	0	Center line
MN2004	3	0	0	0	0	0	0	0	0	0	0	Edge of road

**Figure C 4: “02Distances” opened in Microsoft Access.**

Structuring the database as required facilitated the use of R for later data visualization and statistical analysis—distance and concentration data were extracted separately, and then matched by the unique study identifier.

**Figure C 5: “PM25measurements” opened in Microsoft Access.**